

Access DB# 64900

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval  
 Reference Librarian  
 Biotechnology & Chemical Library  
 CM11 1E07 - 703-308-4488  
 jan.delaval@uspto.gov

\*\*\*\*\*

## STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Jan</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>4448</u>	AA Sequence (#) <u>✓</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>4/22/02</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>4/22/02</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>✓</u>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>4:00</u>	Other _____	Other (specify) _____

This Page Blank (asp.c)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 14.69 seconds

(without alignments)  
2136.414 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099  
Sequence: 1 MAEPLRGSRGSRGSGRRAR.....GKEPLADLGESGLIEYWG 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154.5	7.4	1217	2 T00270	hypothetical prote
2	145.5	6.9	1072	1 A37221	neurofilament trip
3	140	6.7	1634	2 T26517	hypothetical prote
4	137.5	6.6	854	2 S02003	neurofilament trip
5	134.5	6.4	587	1 WMBETE	65k early nonstruc
6	133	6.3	971	2 T19431	hypothetical prote
7	132.5	6.3	792	2 T29187	hypothetical prote
8	132	6.3	586	1 WMBEDE	65k early nonstruc
9	128.5	6.1	635	2 T09648	nucleolin homolog
10	128.5	6.1	734	2 B42680	nucleolin homolog
11	128	6.1	303	1 Z8BP22	gene 8 protein - p
12	127	6.1	1365	2 T14871	suppressor two of
13	126	6.0	856	2 T16543	hypothetical prote
14	124.5	5.9	1684	2 T00057	gravin - human
15	124.5	5.9	1791	2 T02345	hypothetical prote
16	122	5.8	1082	2 T15269	hypothetical prote
17	120	5.7	669	2 T15269	hepatoma-derived g
18	119	5.7	450	2 T08701	hypothetical prote
19	119	5.7	765	2 T49592	neurofilament trip
20	119	5.7	1603	2 S17983	gene posterior sex
21	119	5.7	1611	2 T38236	hypothetical prote
22	118.5	5.6	508	2 T04605	hypothetical prote
23	118	5.6	577	2 T12536	hypothetical prote
24	117.5	5.6	1110	2 T15116	NF-180 - sea lamp
25	117.5	5.6	2187	2 T30826	nascent polypeptid
26	117	5.6	553	2 C75318	hypothetical prote
27	116.5	5.6	714	2 T22454	hypothetical prote
28	116.5	5.6	955	2 S52959	male-specific leth
29	116.5	5.6	974	2 A40580	lodestar maternal-

30	116	5.5	672	2 H86169	hypothetical prote
31	116	5.5	1043	2 A56037	DNA-binding protei
32	115.5	5.5	872	2 S62061	SCD5 protein - yea
33	115.5	5.5	1280	2 T00365	hypothetical prote
34	115	5.5	382	2 T14336	RAD23 protein, iso
35	114	5.5	891	2 G84693	hypothetical prote
36	114	5.4	1274	2 T16251	hypothetical prote
37	113.5	5.4	2361	2 T25752	hypothetical prote
38	113	5.4	611	2 T06458	nucleolin homolog
39	113	5.4	776	2 T20738	hypothetical prote
40	112.5	5.4	95	2 J04760	SWI3 protein - hum
41	112.5	5.4	675	2 T03744	myod protein inhib
42	112.5	5.4	2109	2 T33247	hypothetical prote
43	112.5	5.4	3942	2 T42730	Bassoon protein -
44	112	5.3	480	2 J07552	Shb-like adapter p
45	112	5.3	900	2 C96842	hypothetical prote

## ALIGNMENTS

RESULT 1  
T00270  
hypothetical protein KIAA0596 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #extl\_change 21-Jul-2000  
C:Accession: T00270  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
A:Reference number: 214086; MUID:98290545  
A:Accession: T00270  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1217 <NAG>  
A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA25522.1; PID:g3043716  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0596

Query Match	7.4%	Score 154.5;	DB 2;	Length 1217;
Best local similarity	23.5%	Pred. No. 0.086;		
Matches 105;	Conservative 48;	Mismatches 164;	Indels 129;	Gaps 21;
QY	6	RGGRPSRGRGRARRRARGRCPPRARQSPARLIPVTLVDSDEVEYVADVEYVP	65	
DB	455	RQGRGRQGRSPQASG-----PNRRQAPSMLSPPA---LSSDSKDEGDEGTEELP	506	
QY	66	-----VARLPAPAKPPRODSDEGAEE-----GPAGAPRTLVRRRRRL	106	
DB	507	ALPVLAKSTIKKALASVPSPALPRLPSLHWEKRAQESVGFDPAPAPANPGP---RRGRWV	563	
QY	107	DPEGAPVVPYSGKVOSSLNLIPDNSSLKLPSS--EP-EDPADLTNNGSSSPEDDALPS	163	
DB	564	QPG-----VELSVRSMIDLRLQLETLAPSLQDPQSDSLATIPSGRKKGQALFT	612	
QY	164	G-----SPRRKKLRKK-----EKKE-----KMEFPQODISLPDPS---SRNK	201	
DB	613	SLTQNEKPPRPQASQPCSVPIIRLLSQEGVADLEAPLIEDIIVPEPSDNPMTDT	672	
QY	202	SRKHTALQK-----LREYNKRLOD-----LRSLSPKHOQSPALQSTDEYVU	245	
DB	673	SEQVQAPARAGTIGRYVPSGRSSEKSPSACVYDSSSLSPEN--PREDSESTEPS	730	
QY	246	VEGPVLPOSSRLFTLTKICRADLVRLVRKSEPLQNVVDMANHLGVSPNRILLFGESE	305	
DB	731	VDD-----ISSDLEAPADEDEEEEGGMP-----YGLQE	762	
QY	306	LSPTAP-----SPLKIGVADIIDICVYLASSSEATFSSQELRLVQGRKKQMLE	355	
DB	763	GSP--QTPDQDQFLKHQFETLASGAPGAP--VOYPERSESRSISSRFLQVQTRPREP--	818	

QY 356 ISLSPDPLKLMHSYEAAGLSGRK 381  
 DB 819 ---SPSSSLALMSRPAQVPOASGEQ 841

RESULT 2

A37221  
 neurofilament triplet H protein - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Dec-1999  
 C:Accession: A37221; A25649; A30796; A32757; B25649  
 R:Chin, S.S.M.; Liem, R.K.H.  
 J. Neurosci. 10, 3714-3726, 1990  
 A:Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vim  
 A:Reference number: A37221; MUID:91038277  
 A:Accession: A37221  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1072 <CHT>  
 A:Cross-references: GB:AF031879; NID:q2642597; PIDN:AAB87068.1; PID:q2642598  
 R:Robinson, P.A.; Wion, D.; Anderton, B.H.  
 FEBS Lett. 209, 203-205, 1986  
 A:Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).  
 A:Reference number: A25649; MUID:87080760  
 A:Accession: A25649  
 A:Molecule type: mRNA  
 A:Residues: 230-318:472-542 <ROB>  
 A:Cross-references: GB:M37227  
 R:Dautigny, A.; Pham-Dinh, D.; Rousset, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.  
 Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988  
 A:Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ det  
 A:Reference number: A30796; MUID:88309090  
 A:Accession: A30796  
 A:Molecule type: mRNA  
 A:Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-  
 A:Cross-references: GB:M21964; NID:q205685; PIDN:AAA1695.1; PID:q205686  
 R:Lieberburg, I.; Spiner, N.; Snyder, S.; Anderson, J.; Goligaber, D.; Smulowitz, M.; C  
 Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989  
 A:Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide  
 22.  
 A:Reference number: A32757; MUID:89184647  
 A:Accession: A32757  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <DIE>  
 A:Cross-references: GB:J04517; NID:q205679; PIDN:AAA1692.1; PID:q205680  
 C:Superfamily: neurofilament triplet H protein  
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.9%; Score 145.5; DB 1; Length 1072;  
 Best Local Similarity 22.1%; Pred. No. 0.26;  
 Matches 95; Conservative 63; Mismatches 166; Indels 105; Gaps 20;

QY 4 PLRGPRSRGSGARAGRCPRAROSPAPRL-IPDTVLVDVSDSEV-----55  
 DB 613 PVEAKSPAEKSPASVSKSG-EAKSPAEKSPAEVKSPTAVKSPAEKSPAEKSPVTAK 671  
 QY 56 --LEVADPEVPPARLPAPAKPEODSDSDSEGAEGPAGAPRTLVRRRRRLDPEAV 113  
 DB 672 SPRAKSPPEV-----KSPASVKSPEAKSPAGAKSPAEKSPVAKSPAEKSPAEK 725  
 QY 114 VPYSGKVOSSLILPDNSLKLCPSEPEDEADLTN-----SGSSPSEDDA---LP 162  
 DB 726 -PPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 779  
 QY 163 SGSPWAKKLKKCKEKKKEEPPDDDISPLPPS-SRNKSR---KHTEALOKLEAVK 217  
 DB 780 AKSPEKSPKSVKEIKPPAEVKSPEKSPKAEKSPAEKSPAEKSPAEKSPAEKSPAEK 839  
 QY 248 RLQDLRSCSPKHOHSPALQSTDEVLVEGPVLQSSRLFTLKICGRADIVLP---VR 274  
 DB 840 RPADIR---SPDOVKSPEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 886

QY 275 MSEPLQNVVDHMANHIGVSPNRILLFGESELSPPATPSTLKGADIIICVLAASSSEA 334  
 DB 887 AKEP-----PKKV-----EEKTP-ATPKTE-----VKESKDEA 915

QY 335 TETSOELRLVOGKKEKHOMLEISLSPDPLKVMHSYEAAGLSGHKISFFDGTLSGK 394  
 DB 916 PKEAO-----KPKAEKEPLETE--KPKDSGAEKKEAEKRAAAPE-----E 956

QY 395 ELPADLGL 403  
 DB 957 ETPAKLGAK 965

RESULT 3

T26517  
 hypothetical protein Y18D10A.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26517  
 R:Harris, B.  
 submitted to the EMBL Data Library, December 1998  
 A:Reference number: Z20226  
 A:Accession: T26517  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1634 <WIL>  
 A:Cross-references: EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1  
 C:Genetics:  
 A:Gene: CESP:Y18D10A.1  
 A:Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2

Query Match 6.7%; Score 140; DB 2; Length 1634;  
 Best Local Similarity 22.0%; Pred. No. 0.94; Indels 106; Gaps 18;  
 Matches 88; Conservative 54; Mismatches 152;

QY 3 EPLRGPRSRGSGARAGRCPRAROSPAPRL-IPDTVLVDVSDSEVLEVADP- 61  
 DB 413 EMMKGSYVKEKAAARRAEASAGRSQAPGAPAPA-----ASELQDDPPDGLSMSPDG 465

QY 62 -----VEVPV-----ARLPA-PAKPEODSDSDSGAAEGPAGAPRTLVRRRR 102  
 DB 466 SDSETEHOKQKHIPAMVTRRSARLSALPVPKKASSSK--MPPPSPPSTPGRGR 523

QY 103 R-----RLDGEAPVNVVYSGVOS-SLNLIPNSSLKLCPSPEDEA-DLNSGSS 154  
 DB 524 RRRITLSMMEPAAAVTPAPRGPRSRSAKVSENTEPLSEAPAPVYKRGGRSRST 583

QY 155 PS-EDDALPSGSPWAKKLKKCKEKKMEPPDDDISPLPPSSRNKSRKHT--EALOK 211  
 DB 584 MSLTDESEPTSTSTAKRSKRAESDEEDQDLINKSP-EKPKKPSKTTTETVGDVLK 642

QY 212 -IREVVK-----RLQDLRSCSPKHOHSPALQSTD-----DE 242  
 DB 643 RLRLDAKTATATVIRPGPELRTKMERMRAPAVVSKKEKKNAGSADSSINEEHDE 702

QY 243 VVLVEGPV--LPOSSRLFTLKICRADIVLPVRMSEPLQNVVDHMANHIGVSPRILL 300  
 DB 703 TWILEQTLDLPOQTSODEPRISGSEL-----LDEQFDSSEHSGTVPS----- 747

QY 301 FGESELSPPATPSTLKGADIIICVLAASSSEATSTQ 340  
 DB 748 --APELTKNPAPV-----PEASASAE 768

RESULT 4

S02003  
 neurofilament triplet H protein - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 13-Aug-1999  
 C:Accession: S02003



[illegible]

## RESULT 7

129187  
 hypothetical protein C55C3.3 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: J29187  
 R:Woessne, J.; Steiyyes, L.  
 submitted to the EMBL data library, April 1996  
 A:Description: The sequence of *C. elegans* cosmid C55C3.  
 A:Reference number: Z20585  
 A:Accession: J29187  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1\*792 <MODE>  
 A:Cross-references: EMBL:U053335; PTDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3  
 A:Experimental source: strain Bristol N2; clone C55C3  
 C:Genetics:  
 A:Gene: CESP:C55C3.3  
 A:Map position: 4  
 A:Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 286/2; 327/1; 362/3; 393/3; 43

RESULT 8  
KMBE6E  
65K early nonstructural protein - human cytomegalovirus (strain AD169)  
N:Alternate names: UL84 protein  
C:Species: human cytomegalovirus, human herpesvirus 5  
A:Note: host Homo sapiens (man)



[illegible]

**RESULT 13**

T16543  
hypothetical protein K03C7.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Mar-2000  
C:Accession: T16543  
R:Leimbach, D.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid K03C7.  
A:Reference number: Z18532  
A:Accession: T16543  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-856 <LEI>  
A:Cross-references: EMBL:U40059; NID:g1055170; PID:g1055171; PIDN:AAA81138.1; CESP:K03C7.1  
A:Gene: CESP:K03C7.1  
A:Introns: 176/1; 632/2; 653/1; 717/3; 746/3; 791/3  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

**Query Match**      6.0%; Score 126; DB 2; Length 856;  
Best Local Similarity 21.4%; Pred. No. 3;  
Matches 85; Conservative 55; Mismatches 153; Indels 104; Gaps 17;

**Db**    551 ASGE 554

**Query**    2 AEPILGRGPRSGRGARARGRCPRAROSPARIIPDTVLVDVSDSEVLEVADP 61  
         |::| |::| :  
**Db**       382 ADPL--GGPKKKDKPRLAKKAPAK--PTDKPKK-----DARKDAKPTPEPAKP 427  
         | :  
**Qy**       62 V-----EVPARLPAPAKPEQDSDSDEGAEGPGA----PRTIVR 100  
         | :  
**Db**       428 VAPKWKPPWEEDPDPEPEPADFTMPKKRKPTDEDPA---PLGPNKKDKPRLAKKA 482  
         | :  
**Qy**       101 RRRRLD---PEAAVVVYYSKGVOSSLNLIPDNS-----LLKLCPSEPDEADLTNGS 153  
         ::  
**Db**       483 PTKKPADPKPKEEEEEKFPVAPKWKRPWEEDPDDEDPADFVPIKPGEDEDEPEDADEE 542  
         | :  
**Qy**       154 SPSEDALPSGSPWPKLRKKCEEKKEEPPDDISPLPO----- 196  
         | :  
**Db**       543 PEDE--PADEEPKAKKPKKHKRKKPKPVVEEKEPTPEPVYPKAPKIATIKREE 599  
         | :  
**Qy**       197 -----SSRNKSRAKTEALQKLEVENKRLODLSCSLPKQHSQPAL----- 236  
         | :  
**Db**       600 PIPIPKPEKTIERNKKEERIPALRYAKK-PRELVYIPIVIPME-QTALITOGMGAF 657  
         | :  
**Qy**       237 ---QSTDEVYLVEGPVL--QSRLFTL-----KICRADLYNLPRMSEPLQNVYDH 285  
         ::  
**Db**       658 GKSRANAEVNNGDPRIVOGAVDSKVYIPLWNDESKCANRSQMATAFGA-PRRIDQNVYDH 716  
         | :  
**Qy**       286 MANHLGVSPNRILLFGESELSPFATPSTLKGVADI 322  
         | :  
**Db**       717 ---HKKSGGIPLLAKTGTTHPHGEYGTIRKQTADV 750  
         | :

**RESULT 14**

JM0057  
gravain - human  
C:Species: Homo sapiens (man)  
C>Date: 13-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 08-Oct-1999  
C:Accession: JM0057  
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.  
J. Biochem. 123, 119-1126, 1998  
A>Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial  
A:Reference number: JM0057; MUID:98269042  
A:Accession: JM0057  
A:Molecule type: mRNA  
A:Residues: 1-1684 <SAT>  
A:Cross-references: DBJ:AB003476; NID:g2081606; PIDN:BA119927.1; PID:d1020716; PID:g



F:433-439/Region: nuclear location signal  
 F:522-527/Region: nuclear location signal  
 F:591-596/Region: nuclear location signal  
 F:671-676/Region: nuclear location signal

Query Match 5.9%; Score 124.5; DB 2; Length 1684;  
 Best Local Similarity 20.7%; Pred. No. 8.5;  
 Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

QY 43 VLVDVSDSEEVLEVADPEVPAVRLPAKPEQSDSD----- 82  
 Db 32 VVHDITDGGDETEPEIIEIOPSSSENLLETQPTESQANDIGFKVEVGFKFTVKDK 91  
 QY 83 -----SQAAGPAGA-----PRTLVRRRRLDPGEAPVVPYSGK-VQSS 124  
 Db 92 TEKPDYVOLLTVKKDEGEAGAGDHKDPISLAGEAASKESPEKOSTEKPEETLKREQSH 151  
 QY 125 LNLIP--DNSSLILKICSEPED--EADLTNSGSSPEDDALPSGSP-----WRK 169  
 Db 152 AEISPPAESGQAVECKEKEGEKQEKESKSAESPTSPVISEGTFRKKFTQGMAGMRK 211  
 QY 170 K--LRKCKEK-----EKKMEFPDODISPLPQSSRNKSRKHTALOKLREVNKRLODLR 223  
 Db 212 KTSFRKPEDEVEASEKKKEQEPK-----VDTEEDGKAEVASEKLTASEQAHQ-EPAE 265  
 QY 224 SCLSPK---OHOSPALOSTDEEVLVVEGP---VLPOSSRLFTLKIRGRADVLRLPYRMS 276  
 Db 266 SAHPRLSAEYKVELPS-EEQVSGSGSPSEKPAPLATEYFDEKIEYHOEEVVAEVAHV 324  
 QY 277 EPLQNVYDHMANHLGVSPNRIILLFGESELSPTATPTLKLGVADIIDCVVLASSSEATE 336  
 Db 325 TVEERTEEQ-----KTEVEETA-----GSVPAEELVEMDAEPQEA 360  
 QY 337 TSQEL-RLR---VOGKEKHQMLEISLSPDPL-----KVLASHYEPAGLSGHKL 382  
 Db 361 PAKELVLIKETCVSGEPTQAD--LSPDEKVLKRPPEGVVEVEMLSQERMKVQGSPL 418  
 QY 383 SFFPDGT---KLSGKELPADLG---LESGDLIEV 410  
 Db 419 KKLFTSTGLKLSGKKOKGKRGGDSESGHTQV 452

## RESULT 15

T02345  
 hypothetical protein KIAA0324 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999

C:Accession: T02345

R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;

re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.

submitted to the EMBL Data Library, March 1998

A:Description: Sequencing of human chromosome 16p13.3.

A:Reference number: Z14664

A:Accession: T02345

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1791 <R10>

A:Cross-references: EMBL:AC004493; NID:q2996648; PIDN:AA08453.1; PID:q2996650

C:Genetics:

A:Map position: 16

A:Introns: 1610/2; 1706/2

A>Note: KIAA0324

Query Match 5.9%; Score 124.5; DB 2; Length 1791;  
 Best Local Similarity 25.3%; Pred. No. 9.2;

Matches 72; Conservative 26; Mismatches 96; Indels 91; Gaps 13;

QY 34 SPARLIPDVLVLDVSDSEEVLEVADPEVPAVRLPAKPEQSDSDSEGAAGPAGA 93  
 Db 414 SSELSPDAVEKAGWSSNQ-----SISSPVLDAVPRPSRER-----SSASSPKMGDL 463

QY 94 PRTLVRRRRLDLPGEAPVVPYSGKVOSSLNIPNSSL-----LKLCPSEPD 144  
 Db 464 PRTSRRRS-----SGSSPGLRDSG-----TPSRHLSGSSPGMKDIPRPSRGRS 510  
 QY 145 EADLTNSGSSPEDDALP-----SGSPWRKRLKCK-----EKEKKMEFPDQ----- 188  
 Db 511 ECD-----SSP-EPKALPQTPRPSRSPSPSELNNKCLTPQRRSGSSSVQDKTARTP 564  
 QY 189 -----DISLPQPSR-----MKSRKHTALOKLREVNKRLODLRSCLS 227  
 Db 565 LGQSRSGSSQELDVKPSASPOERSSESDSDPSAKTRTPLQRSGSSPEVDSKSRLS 624  
 QY 228 PKQHOS-----PALOSTDEEVLVVEGP---VLPOSSR 256  
 Db 625 PRSRSGSSPEVKDKPRAAPRAQSGSDSSPEPKAPAPRALPRSR 669

Search completed: April 22, 2002, 10:07:02  
 Job time: 47 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 13.21 Seconds  
(without alignments)  
1143.521 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099  
Sequence: 1 MAEPLRGGRGRRGRRGRR.....GKELPADLGESGLIEWWG 412

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.5	6.5	831	1	NFH_RAT
2	134.5	6.4	587	1	U844_HCMVT
3	132	6.3	586	1	U844_HCMVA
4	128.5	6.1	704	1	NP14_RAT
5	128	6.1	303	1	VG08_BP22
6	127	6.1	1365	1	SU22_DROME
7	122.5	5.8	1781	1	AKAC_HUMAN
8	121	5.8	1109	1	TCF8_RAT
9	119	5.7	1603	1	PSC_DROME
10	116.5	5.6	1039	1	MSL1_DROME
11	116.5	5.6	1061	1	LDS_DROME
12	116	5.5	1043	1	TCF8_MESAU
13	115.5	5.5	872	1	SCD5_YEAST
14	114.5	5.5	1411	1	TCOF_HUMAN
15	113.5	5.4	1280	1	DYNA_RAT
16	112.5	5.4	95	1	SM32_HUMAN
17	112	5.3	919	1	DNL1_HUMAN
18	112	5.3	943	1	CENC_HUMAN
19	112	5.3	1004	1	PRPX_RAT
20	111.5	5.3	1170	1	XPE_MOUSE
21	111	5.3	741	1	BSG2_DROME
22	111	5.3	897	1	EP15_MOUSE
23	111	5.3	1281	1	DYNA_MOUSE
24	110.5	5.3	919	1	SYND_HUMAN
25	110.5	5.3	2517	1	NCR2_HUMAN
26	110	5.2	542	1	TUL1_HUMAN
27	110	5.2	612	1	ARRS_MAIZE
28	110	5.2	883	1	E74B_DROME
29	110	5.2	1206	1	FM14_MOUSE
30	110	5.2	2150	1	SDC3_CAEEL
31	109.5	5.2	536	1	GAC_MVCB
32	109.5	5.2	972	1	ORF4_SCHRO
33	109	5.2	974	1	YMB4_CAEEL

34	109	5.2	1089	1	Y553_HUMAN	O9ukf3 homo sapien
35	108.5	5.2	487	1	ATP2_CHICK	O93602 gallus gall
36	108.5	5.2	505	1	TUB_MOUSE	P50586 mus musculu
37	108.5	5.2	1117	1	TCF8_MOUSE	O64318 mus musculu
38	108	5.1	487	1	ATP2_HUMAN	P15336 homo sapien
39	108	5.1	487	1	ATP2_MOUSE	P16951 mus musculu
40	108	5.1	673	1	FXR2_HUMAN	P51116 homo sapien
41	108	5.1	1402	1	IFG4_RABIT	P41110 cryctolagus
42	107.5	5.1	532	1	ICP0_HSVB	P28890 equine herp
43	107	5.1	487	1	ATP2_RAT	O00969 rattus norv
44	107	5.1	650	1	MDCL_XENLA	P20397 xenopus lae
45	107	5.1	1224	1	DYNA_CHICK	P35458 gallus gall

#### ALIGNMENTS

```

RESULT 1
ID      NFH_RAT
AC      P16884: 063368;          STANDARD;          PRT;      831 AA.
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE      (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
GN      NEFH OR NFH.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=89065087; PubMed=3143606;
RA      Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
RT      "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT      Identification of putative phosphorylation sites."
RL      FEBS Lett. 241:213-218(1988).
RN      [2]
RP      SEQUENCE OF 37-831 FROM N.A.
RX      MEDLINE=88309090; PubMed=2457365;
RA      Daughy A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
RA      Jolles P.;
RT      "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT      in situ detection."
RL      Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RN      [3]
RP      SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX      MEDLINE=87080760; PubMed=2878828;
RA      Robinson P.A., Wion D., Anderson B.H.;
RT      "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT      (NF-H)."
RL      FEBS Lett. 209:203-205(1986).
RN      [4]
RP      SEQUENCE OF 318-831 FROM N.A.
RX      MEDLINE=89184647; PubMed=2928342;
RA      Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,
RA      Smolowitz M., Carroli Z., Emanuel B.S., Bretiner J., Rubin L.;
RT      "Cloning of a cDNA encoding the rat high molecular weight
RT      neurofilament peptide (NF-H): developmental and tissue expression in
RT      the rat, and mapping of its human homologue to chromosomes 1 and
RT      22."
RL      Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
-1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
-1- AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
-1- NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
-1- SUBSERVED BY THE TWO SMALLER NF PROTEINS.
-1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P, NFH IS
-1- PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
-1- THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
-1- INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
-1- OF AXONAL CALIBER.

```

```

-1- PMM PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGE NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M37227; AAA41693.1; ALT_FRAME.
DR EMBL; X13804; CA32038.1; ALT_FRAME.
DR EMBL; M21964; AAA41695.1;
DR EMBL; J04517; AAA41692.1; -.
DR PIR; A30796; A30796.
DR PIR; A25649; A25649.
DR PIR; B25649; B25649.
DR PIR; S02003; S02003.
DR InterPro; IPR001664; IP.
DR Pfam; PF00038; filament_1.
DR ProSite; PS00226; IF.1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
FT NON_TER 1
FT DOMAIN 1
FT FT 276 641 51 x 3 AA TANDEN REPEATS OF K-S-P.
FT CONFLICT 164 164 L -> I (IN REF. 2).
FT CONFLICT 185 185 I -> S (IN REF. 2).
FT CONFLICT 193 193 L -> T (IN REF. 2).
FT CONFLICT 199 199 M -> T (IN REF. 2).
FT CONFLICT 346 346 K -> N (IN REF. 1).
FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).
FT CONFLICT 485 485 P -> S (IN REF. 2).
FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A -> V (IN REF. 4).
FT CONFLICT 757 759 AAP -> GST (IN REF. 4).
FT CONFLICT 769 769 T -> L (IN REF. 2).
FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

Query Match 6.5%; Score 135.5; DB 1; Length 831;
Best Local Similarity 21.9%; Pred. No. 0.377;
Matches 94; Conservative 61; Mismatches 169; Indels 105; Gaps 20.

4 PLGRGPRGRGRCRGARARGRCPRARQSPARL-IPDTVVDLVSDDEY----- 55
372 PAEAKSPAAKSPAAKSGC-EAKSPAEEKSPADEVKSPATVSPVAAKPAEYKSPYTK 430
56 --LEVADPEVVARLPADAPPEODSDSEGAEGCAPPTLVRRRRRLIDGEADV 113
431 SPAEKSPVEV-----KSPAYSKSPSEKSPACAKSPAAPVAAKSPAEEKSPAGAK- 484
114 VVYVSGKQSSLNLIPDNSLLKLCSEPEDEDADLTN-----SGSSPSEDDA---LP 162
485 -PRAEKSPAEEKSPAEEKS-----PAAKSPAEEKSPVEVKSPEKAKSPVEKAGKSLAE 538
163 SGSPRRKTKRKCKEKEEKMEEPDODIAPL-POPSRRKSR-----KHEEALOKLEVNK 217
539 AKSPEKAKSPVDEIKPPAEVKSPEKAKSPMKREKAKSPKATLDVYKSPAEKPAKEAK 598
218 RLQDLRSLCSLPRKQHSALQSTDEEVVLEGPVLPQSSSLFTLKTRCRADLVRLP---VR 274
599 RADIR---SPQVKSPPAEKAKS-----PEKEETPREKVAPKKEEVKSPVEEVK 645
275 MSEPLQNVVDHMANHLGVSPNKLILFGSELSLPTATPTSLKLGAVADIIDCVYLAASSSA 334

```

DB	AKEP	PKKV	EEKTP	ATPKTE	VKSKRDEA	674
QY	335	TETSEELRLRQGGKKHQMELISLSPDPLKVLMSHYEAMGLSCHKLSFFPDGKLSGK	394			
Db	675	PKRAE-----KPKAEKEPLTE--KPKDSPEGAKKEBEAKKNAAPE-----E	715			
QY	395	ELPADIGLE	403			
Db	716	ETPAKLGK	724			
RESULT 2						
ID	UL84_HCMWT	STANDARD:	PRT:	587	AA.	
AC	P29839:					
DT	01-APR-1993 (Rel. 25, Created)					
DT	01-APR-1993 (Rel. 25, Last sequence update)					
DT	01-APR-1993 (Rel. 25, Last annotation update)					
DE	65 KDA EARLY NONSTRUCTURAL PROTEIN (UL84 PROTEIN).					
GN	UL84.					
OS	Human cytomegalovirus (strain Towne).					
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;					
OC	Betaherpesvirinae; Cytomegalovirus.					
OX	NCBI_TaxID=10363;					
RP	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=92114132; PubMed=1309892;					
RA	He Y.S., Xu L., Huang E.S.;					
RT	"Characterization of human cytomegalovirus UL84 early gene and					
RT	J. Virol. 66:1098-1108(1992)."					
RL	J. Virol. 66:1098-1108(1992).					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation --					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC	-----					
DR	EMBL; M81432; AAA45947.1; -.					
DR	PIR; A41808; WMBETE.					
KW	Nonstructural protein.					
FT	DOMAIN 9 19 ARG-RICH (BASIC).					
FT	DOMAIN 162 170 LYS-RICH (BASIC).					
FT	DOMAIN 171 183 ASP/GLU-RICH (ACIDIC).					
SO	SEQUENCE 587 AA: 65388 MW: 130170E41FB32208 CRC64:					
Query Match 6.4%; Score 134.5; DB 1; Length 587;						
Best Local Similarity 24.4%; Pred. 0.28;						
Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;						
QY	6	RGGRPRSR---GNGARRAGARGRCPRAROS---PARLPTVLVDL--VSDSEEV	55			
Db	11	RARRRRPARGGGGCVGSNSRHS-GCKRQRRALAPPLTFIATTTTMMGVASTDD	69			
QY	56	LEVADEVPEVVARLAPAKPEQDDSDSGAEGPAGARTVRRRRRLDPGAPVP	115			
Db	70	LLTKLPDEL-----DKSGSQITL-----TLDDKDIRQPR	101			
QY	116	VYSGR---VOSSLNLIPDN-----SLKLICPSEPEADLTNGSSPSSEDALPS	163			
Db	102	VARGVYHLIQLHLDRPELRDPFQILLSTPLQLEANGESQ--TAPATSGEETA--	156			
QY	164	GSPMKRLKLRKCKEKKKEEFPDODISFLPQSSRNKSRKTTALQKLRVNNKRLQDR	223			
Db	157	---SHELEKKEKKEKKEE--DEB-----DRNDRER---GILCVSNSDSDVR	198			
QY	224	SLTS--PKHOSPALOSTDEVVLVGEVPLTPQSSRLFTLKICRADYLRVPMSEPLQN	281			
Db	199	PAFSLPAPAPGCHILASVYDQ-QLTRMAIVRLSLNLFALRI--ITPPLKRVPLRKAHAHT	256			

[illegible]

RESULT	3				
ID	UL84_HCMVA	STANDARD:	PRT:	586 AA.	
AC	UL84_HCMVA				
AD	PI6727;				
DT	01-AUG-1990 (Rel. 15, Created)				
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	01-APR-1993 (Rel. 25, Last annotation update)				
DE	65 KDA EARLY NONSTRUCTURAL PROTEIN (UL84 PROTEIN).				
GN	UL84.				
OS	Human cytomegalovirus (strain AD169).				
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;				
OC	Betaherpesvirinae; Cytomegalovirus.				
OX	NCBI_TaxID=10360;				
RN	11				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90269039; PubMed=2161319;				
RA	Chee M.S., Bankier A.T., Beck S., Bohm R., Brown C.M., Cerny R.,				
RA	Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,				
RA	Predede E., Satchwell S.C., Tomlinson P., Weston K.M., Barrall B.G.;				
RT	"Analysis of the protein-coding content of the sequence of human				
RT	cytomegalovirus strain AD169."				
RL	Curr. Top. Microbiol. Immunol. 154:125-169(1990).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; X17403; CAA35358.1; -				
DR	PIR; S09848; WMBEDE.				
KM	Nonstructural protein.				
FT	DOMAIN 9 19				ARG-RICH (BASIC).
FT	DOMAIN 162 170				LYS-RICH (BASIC).
FT	DOMAIN 171 182				ASP/GLU-RICH (ACIDIC).
SO	SEQUENCE 586 AA; 65428 MW; 54AB9120b077223f CAC64;				

Query Match	6.3%	Score 132	DB 1	Length 586
Best Local Similarity	23.6%	Pred. No. 0.39		
Matches 104	Conservative 62	Mismatches 145	Indels 130	Gaps 25

[illegible]

OY	224	SCLS--	PKOHSPALQSTDDDEVVLVEGCVLPQSSRLTTLKIRCAADLVLPVKNSEPLON	281
Db	198	PAFSFPPRPSCILIRSYIDQ	LTIRMAIVLSINLEALRI--TTPLLKRIPLR----	RK 250
OY	282	VVDHANILGVSPRIILTEGESELSPTAPSTLKLGVADIIDCVLASSSEATETSOEL		341
Db	251	AAHHHTALH----	DCLAHLPRLFFEPRLDINNTENNAASVAD---TAESTDA-DLTPTL	301
OY	342	RLRVQGEKKHOMLEISLSPDSPKLVLMSHYEANGLSG-----	HKLS-----	363
Db	302	TVRVV-----	HALCWHRYEG-GISGPRGLTSRISARLESETTAKTLP	342
OY	384	FFPDGTKLSKEIPLADIGLES		404
Db	343	SVFGRLDIPNESPPDLTLLS		363

	RESULT	4
NP14_RAT		
ID	NP14_RAT	STANDARD; PRT; 704 AA.
AC	P41777;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	01-NOV-1995 (Rel. 32, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	NUCLEOLAR PHOSPHOPROTEIN P130 (NUCLEOLAR 130 KDA PROTEIN) (140 KDA NUCLEOLAR PHOSPHOPROTEIN) (NOPP140).	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Rattus.	
OX	NCBI_Taxid=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 292-309 AND 563-601.	

RA Meier U.B., Blobel G.:  
RT "NOPp140 shuttles on tracks between nucleolus and cytoplasm.",  
RL Cell 70:127-138(1992).  
CC  
CC -1- FUNCTION: RELATED TO NUCLEOGENESIS, MAY PLAY A ROLE IN THE  
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER  
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC  
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN  
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).  
CC  
CC -1- SUBUNIT: INTERACTS WITH DKC1/NAP57.  
CC  
CC -1- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN  
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE  
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOPLASM TO  
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.  
CC  
CC -1- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND  
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPp140 IS ONE OF THE  
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; M94287; AAAA1718.1; -	Repeat; GMP-binding; ATP-binding
DR	EMBL; M94288; AAAA1719.1; -	
KW	Nuclear protein; Phosphorylation;	
FT	DOMAIN	11 x 12 AA APPROXIMATE REPEATS OF AN
FT	84	570
FT	REPEAT	ACIDIC SERINE CLUSTER.
FT	84	95
FT	REPEAT	ACIDIC SERINE CLUSTER 1.
FT	127	138
FT	REPEAT	ACIDIC SERINE CLUSTER 2.
FT	170	181
FT	REPEAT	ACIDIC SERINE CLUSTER 3.
FT	231	242
FT	REPEAT	ACIDIC SERINE CLUSTER 4.
FT	274	285
FT	REPEAT	ACIDIC SERINE CLUSTER 5.
FT	335	346
FT	REPEAT	ACIDIC SERINE CLUSTER 6.
FT	373	384
FT	REPEAT	ACIDIC SERINE CLUSTER 7.
FT	434	445
FT	REPEAT	ACIDIC SERINE CLUSTER 8.
FT	479	490
FT	REPEAT	ACIDIC SERINE CLUSTER 9.

FT REPEAT 524 535 ACIDIC SERINE CLUSTER 10.  
 FT REPEAT 559 570 ACIDIC SERINE CLUSTER 11.  
 FT MOD\_RES 567 567 PHOSPHORYLATION (BY CK2).  
 FT VARIANT 150 150 MISSING (IN NOP140B).  
 SO SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE4B9EAA3 CRC64;

Query Match 6.1%; Score 128.5; DB 1; Length 704;  
 Best Local Similarity 23.2%; Pred. No. 0.8;  
 Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

QY 33 QSPALIDPTVLDVSDSE-----VLEVADPYEVYARL---PAPAKPEQSDSDSE 84  
 DB 317 QSPKAAAOPTOPADSSADSESDSSSEEEKTPAKTYVSKTPAPAPYKKAESSSDS 376  
 QY 85 GAEGPAGAPRTLVRRRRRLDPCGAPVYV-----YSGKVQSLNL 127  
 DB 377 DSDSDEDPAPRYATNSPLSKPAVTPKPPAKAAVATPKOPAGSGQKQSKADSSSE 436  
 QY 128 IPDNSS-----LKLCPSEPEDEA-----DLTNSGSSPSDED----- 159  
 DB 437 EESSSESEETATKSYTPKARYTAKAAPSILPAKQAPRAGGSDSSSESSSEEEKTPPK 496  
 QY 160 -----ALPSGSPMKKRLK-----KCEKEKKMEFPDDISPLQ----- 195  
 DB 497 PPAKKAAGAAVPPKPPVKKAAESSSSSSSESDSSSEEEKK-----PSKATPKQAGKA 552  
 QY 196 ---PSSRN-KSKRHTALOKLREVNKRLQDLRSCLSPKQHOHPALQSTDEVLVEGPVL 251  
 DB 553 NGVPAQONKAKKKESEEEDEDTQONKKAAGTKPGSGKKRKHNN-----ETPDEA-----AT 602  
 QY 252 POSRRL 257  
 DB 603 POSKRV 608

RESULT 5  
 VG08\_BPP22 STANDARD; PRT; 303 AA.  
 AC P26748;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SCAFFOLDING PROTEIN (PROTEIN GP8).  
 GN 8.  
 OS Bacteriophage P22.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.  
 OX NCBI\_TaxID=10754;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
 RX MEDLINE=91306435; PubMed=1853558;  
 RA Eppler K., Wyckoff E., Goates J., Parr R., Casjens S.;  
 RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging";  
 RT Virology 183:519-538(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kropinski A.M.B., VanderByl C.S.;  
 RT "The completed sequence of genome of Salmonella phage P22";  
 RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 RL -1- FUNCTION: REQUIRED FOR SUCCESSFUL CONDENSATION OF DNA WITHIN THE  
 CC CAVID. THE INTERIOR OF THE PROHEAD IS FILLED WITH THE GP8  
 CC PROTEIN. THE SCAFFOLDING PROTEIN IS LOST FROM THE STRUCTURE  
 CC DURING PACKAGING.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- INDUCTION: THE SCAFFOLDING PROTEIN NEGATIVELY REGULATES ITS OWN  
 CC SYNTHESIS WHEN IT IS NOT ASSEMBLED INTO PROHEADS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL: M59749; AAA72962.1;  
 CC DR EMBL: AF217253; AAF75046.1;  
 CC DR PIR: D40474; Z8BP22.  
 CC KW late protein.  
 SO SEQUENCE 303 AA; 33564 MW; B86BD625D5011ABE CRC64;

Query Match 6.1%; Score 128; DB 1; Length 303;  
 Best Local Similarity 21.3%; Pred. No. 0.31;  
 Matches 73; Conservative 49; Mismatches 100; Indels 120; Gaps 15;

QY 41 DTIVLDVSDS--DEVLVADPYEVYARLPAPAKPEQSDSDSEGAEGPAGAPRTLV 98  
 DB 25 DSLVYDANNDNAGOEGEYV-----LKDDEYAPKQDPAPAKNAEFARRRI 69  
 QY 99 RRRRRRLDPCGAPVYVYSGKVQSLNLIPDNSSLKLCSEPEDEADLTNSGSSPED 158  
 DB 70 RRRQRELEQOMEA---VKRGELPESLRYNDP-----LPPQDINLYLSEGLAKYDY 118  
 QY 159 D---AL---PSSGSPMKKRLK---KCEKEKKMEFPDDISPLPQSSRNKSRKHT 207  
 DB 119 DNSRLALAFNANMTWMLKADARSNAAVEGRKQETQOQAQV-----EAKRHYD 172  
 QY 208 ALQKLRVKNRLQDLRSCLSPKQHOHPALQSTDEVLVEGPVLQSSRLFTLKIRCAD 267  
 DB 173 AAELK-----NIPYQEKEDAFMQLVPPAV-----GAD 200  
 QY 268 LVRLVRKSEPLQNVVDHMANHLGVSPNR---ILLPGESELSPTATPSTLKGVDIID 324  
 DB 201 IMRLPEKSAALM-----YHLGANPEKAROLALMDGQSAI----- 235  
 QY 325 CVVLASSEATETSOELRLRYGKEKHOMLEISLSP--DSP 364  
 DB 236 -----IEIPLSERLTLKPRGK-----QISSAPPADQPI 264

RESULT 6  
 SUZ2\_DROME STANDARD; PRT; 1365 AA.  
 AC P25172;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SUPPRESSOR 2 OF ZESTY PROTEIN (PROTEIN POSTERIOR SEX COMBS).  
 GN SU(2).  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CANTON-S.  
 RX MEDLINE=91279476; PubMed=2057369;  
 RA Brunk B.P., Adler P.N.;  
 RT "The sequence of the Drosophila regulatory gene Suppressor two of  
 RT zeste";  
 RT Nucleic Acids Res. 19:3149-3149(1991).  
 RL -1- FUNCTION: REGULATES EXPRESSION OF THE HOMEOTIC SELECTOR GENES BY  
 CC INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION  
 CC WITH OTHER PROTEINS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: X56798; CAA40134.1;
DR EMBL: X56799; CAA40135.1;
DR PIR: S16845; S16845.
DR PIR: S14871; S14871.
DR Flybase: FBgn008654; Su(z)2.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; zfc3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT ZN-FING 35 73
FT DOMAIN 623 628 POLY-GLN.
FT DOMAIN 1077 1096 POLY-ASN.
FT DOMAIN 1241 1251 POLY-SER.
FT CONFLICT 603 603 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 785 785 K -> N (IN REF. 1; CAA40134).
FT CONFLICT 831 831 A -> R (IN REF. 1; CAA40134).
FT CONFLICT 965 965 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 1065 1065 D -> E (IN REF. 1; CAA40134).
FT CONFLICT 1076 1076 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 1287 1287 A -> P (IN REF. 1; CAA40134).
SQ SEQUENCE 1365 AA; 146058 MW; 7B4B0F35B0FA683 CRC64;

```

Query Match 6.1%; Score 127; DB 1; Length 1365;

Best Local Similarity 23.1%; Pred. No. 2.2; Mismatches 125; Indels 108; Gaps 16;

```

Matches 84; Conservative 47; Mismatches 125; Indels 108; Gaps 16;

QY 110 EAPVVPVYSGKVOSSLNLIPDNSSLLKICPS-EPEDADLTJNSGSSP--SED--DALPSG 164
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 220 ESPMAFCYRLLTYLNDQTKNDENRLSRINODLEPHSVRSKSAKSVTFADLESEIDSG 279
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 165 SPWRKRLRKCKEKEKMEFPDQDISPLPQSSRNK-----SKKHTEA-----LQKRE 214
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 280 SP-RSIVRCK-----TPPVSPSKSKKRLTSSKREAPSPVSNFSLRS 323
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 215 VNRKLQDL-----RSCLSKPOHSPALQSTDEVLEVGVLPOS--SRLETL 260
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 324 NDMRYSDVAVSKYKSPEDQEQFLPREEQQLPEANTNIVYSLPQSLKRSYDADDEL 383
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 261 KTRCADLVRP-----VRMSEPL-----QNVYDHMANHLGV 292
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 384 KLANRKGVGHFLPKLKIELNSMKSLMPLSAGRLDQTSKSSCSAQQDLLEYAANKIGL 443
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 293 SPNRILLLEG-----ESELSPATP-----STLKGVAIIIDCVLASSSEATENS 338
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 444 KPLEQPLQOASNPDSKYSPPNASSPSSSSSTNGSSSSSLGTAD-----ASTSTSTSS 496
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 339 QETRLR-----VQKKEKHQMLEITSLSPDPLKVLAMSHYEAMGLSGHKLSFFPDGTR 390
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 497 HRRKKRKHSHKPRDANGKRRKKLHAELISSQTDGKMKVKIT-----AKPNKLDLFRKSHSL 550
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 391 LSGK 394
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 551 ASGE 554
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 7
AKAC_HUMAN STANDARD; PRT; 1781 AA.
AC Q02952; Q09970; O00498; O00310;
DT 01-FEB-1994 (Rel. 28; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP
DE 250) (MYASTHENIA GRAVIS AUTOANTIGEN GRAVIN).
GN AKAP12 OR AKAP250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Heart;
RX MEDLINE=97153077; PubMed=9000000;
RA Nauert J.B., Klauk T.M., Langeberg L.K., Scott J.D.;
RT "Gravin, an autoantigen recognized by serum from myasthenia gravis
RL patients, is a kinase scaffold protein.";
RL Curr. Biol. 7:52-62(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=98269042; PubMed=9604001;
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT "Changes of gene expression by lysophosphatidylcholine in vascular
RL endothelial cells: 12 up-regulated distinct genes including 5 cell
RL growth-related, 3 thrombosis-related, and 4 others.";
RL J. Biochem. 123:1119-1126(1998).
RN [3]
RP SEQUENCE OF 43-1781 FROM N.A.
RC TISSUE=umbilical vein endothelial cells;
RA Bowditch R.D., Ginsberg M.H.;
RT "Sequence of gravin cDNA isolated from a human umbilical vein
RL endothelial cell library.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1477-1781 FROM N.A.
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=92395179; PubMed=1522245;
RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA Lindstrom J., Ginsberg M.H.;
RT "Molecular cloning and preliminary characterization of a novel
RL cytoplasmic antigen recognized by myasthenia gravis sera.";
RL J. Clin. Invest. 90:992-999(1992).
CC -1- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
CC (PKC).
CC -1- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
CC FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,
CC MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.
CC -1- INDICATION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).
CC -1- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC INVOLVED IN BINDING PKC.
CC -1- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U81607; AAC51366.1;
CC EMBL: AF001504; AAB58838.1;
CC EMBL: AB003476; BAA19927.1;
CC DR EMBL: M96322; AAA35931.1;
CC DR PIR: A43922; A43922.
CC MIM: 604698;
CC InterPro: IPR001573; Pkin_anch.
KW Antigen; Alternative splicing.
FT DOMAIN 603 633 AKAP 1.
FT DOMAIN 752 782 AKAP 2.
FT DOMAIN 827 827 AKAP 3.
FT DOMAIN 98 101 POLY-GLU.
FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).
FT VARSPLIC 1 98 MISSING (IN ISOFORM 2).
FT VARSPLIC 99 106 EEEIVTE -> MLGITTT (IN ISOFORM 2).

```

FT CONFLICT 117 117 E -> K (IN REF. 2).  
FT CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).  
FT CONFLICT 215 215 Q -> K (IN REF. 2).  
FT CONFLICT 448 448 R -> G (IN REF. 2 AND 3).  
FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).  
FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).  
FT CONFLICT 867 867 S -> A (IN REF. 2 AND 3).  
FT CONFLICT 986 986 E -> EE (IN REF. 3 AND 4).  
FT CONFLICT 1530 1530 E -> EE (IN REF. 3 AND 4).  
FT CONFLICT 1581 1581 V -> M (IN REF. 4).  
FT CONFLICT 1601 1601 Q -> L (IN REF. 2).  
SQ SEQUENCE 1781 AA; 191439 MW; BA813937379FAC0F CRC64;

Query Match 5.8%; Score 122.5; DB 1; Length 1781;  
Best Local Similarity 22.0%; Pred. No. 5.7;  
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

OY 40 PPTV-LVLDVSSDDEVLEVADPEVPAVARLAPAK---PEDSDSDSGAAGFAGAPR 95  
DB 192 PPTVOLLTVKKKBGAGAGAGHODPSLAGEAASESEPKOSTEKPEP----- 240  
OY 96 TLVRRRRRLDPGEAPVVPVYSGKVOSSLNLPDSSSLKLCPSPEPD--EADLTNSGS 153  
DB 241 TLKRGSHAEISP-----PAESGA-----VECKEKGEGEKQEKPSKSAE 281  
OY 154 SPSEDDALPSSGP-----WRKK--LRKKCKEKE--EKKMEFPDODISPLPOP 196  
DB 282 SPSPVSTSETGTFKKFTQGWAGMKRKTSPFKPKEDVEASEKKKEQPERK----VDT 336  
OY 197 SSRNKSRTKTELQKLRVKNRKLQDLRCLSPK---QHSPALQSDDDVYLVEGP----- 249  
DB 337 EEDGAEVAASEKLTJASEQHPO-EPESAHEPRLSAEYKVELPS-EECVSSGSGSESK 394  
OY 250 VLPOSSRLFTLTKRCRADVLRLPVMSEPLQNVVDHMANHLGVSPKILLTGESELSPT 309  
DB 395 PAPLTFEVEDELTVEHQEVAEVAHVSTVEERTEEQ-----KTEVEET 437  
OY 310 A--TPSTKLGVADITIDCVVLASSSEATFSTQELRLR---VOGKEKHQMLETSLSPDSE 364  
DB 438 ACSVPAEELVGM-----AEQGEAPAKELVLKTECVSGEDPTQAD--LSPDEKV 487  
OY 365 -----KYLMSHYERAMGLSGHKLSFFPDGT---KLSEKELPADLG---LESGLDI 408  
DB 488 LSKPPEGVVSEVEMLSQDERMKVQSGSPILKLTSTGLKRLSGKKOKGKRGGDESGEHT 547  
OY 409 EV 410  
DB 548 QY 549

RESULT 8  
TCF8\_RAT STANDARD; PRT; 1109 AA.  
AC 062947: 062948: (Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE TRANSCRIPTION FACTOR 8 (ZINC FINGER HOMEDOMAIN ENHANCER-BINDING PROTEIN) (ZFHP).  
GN TCF8.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE OF 6-1109 FROM N.A. AND ALTERNATIVE SPLICING.  
RX MEDLINE=96365389; PubMed=8769566;  
RA Cabanillas A.M.; Darling D.S.;  
RT "Alternative splicing gives rise to two isoforms of zfhpf, a zinc finger/homeodomain protein that binds T3-response elements.";  
RL DNJ Cell Biol. 15:643-651(1996).  
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN

CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/ZFHP-1 (SHOWN HERE) AND  
CC 2/ZFHP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC  
CC FINGER/HOMEDOMAIN PROTEINS.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use, by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: U51583; AAB17130.1; -.  
DR EMBL: U51584; AAB17131.1; -.  
DR HSPB; P08046; 1AIG.  
DR InterPro: IPR000010; Cystatin.  
DR InterPro: IPR001356; Homeobox.  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00031; cystatin; 1.  
DR Pfam: PF00046; homeobox; 1.  
DR Pfam: PF00096; zF-C2H2; 7.  
DR SMART; SM00369; HOX; 1.  
DR SMART; SM00355; Znf\_C2H2; 7.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
KW Transcription regulation; DNA-binding; Nuclear protein; zinc-finger;  
KW Homeobox; Repressor; Activator; Metal-binding; Repeat;  
KW Alternative splicing.  
FT ZN\_FING 150 173 C2H2-TYPE.  
FT ZN\_FING 180 202 C2H2-TYPE.  
FT ZN\_FING 220 242 C2H2-TYPE.  
FT ZN\_FING 248 272 C2H2-TYPE (ATYPICAL).  
FT DNA\_BIND 559 618 HOMEBOX-LIKE.  
FT ZN\_FING 881 903 C2H2-TYPE.  
FT ZN\_FING 909 931 C2H2-TYPE.  
FT ZN\_FING 937 958 C2H2-TYPE (ATYPICAL).  
FT DOMAIN 968 1109 GLU-RICH (ACIDIC).  
FT VARSPPLIC 1 198 MISSING (IN ISOFORM 2).  
SQ SEQUENCE 1109 AA; 121626 MW; BEFE291C8795DDA6 CRC64;

Query Match 5.8%; Score 121; DB 1; Length 1109;  
Best Local Similarity 21.4%; Pred. No. 4;  
Matches 87; Conservative 57; Mismatches 152; Indels 110; Gaps 19;

OY 61 PVEVPVARLPAPAKPEODSDSGAAGPAGAPRTLVRRRRRLDPGEAPVVPVY-G 119  
DB 535 PAQPP-----PAPATEKPESSASSAGNGD-----LSPSOPPLKMLSL 574  
OY 120 KVOSSLNLPDSSSLKLCPSF--PEDEA---DLTNGSSPSSEDDALPSSGPMKKRLR 173  
DB 575 KAYVILNMQPSTEEELTKIADSVNLPLDVYKKWFKEMQAGLITGQSLPEPPSGGNTIPA 634  
OY 174 KCEKEKKME-FEPDOD-----ISP-LPOSSSRNKSRTKTELQKLRVKNRLO 220  
DB 635 KTEEGPQPVYDGNRPQEDSTRGOSPLKMTSSVLPVPGSALNRSRSTSSPSPLNLSANP 694  
OY 221 DLRSCLSPKQHOSPALQSTDEEVILVEGFLPOS--SRIF-----TLKIRCRADL 268  
DB 695 QGVSCVSEGTQEEQVEPLDLTLPRQGGELLRRSTVSSVYQNSVQEEPLNLSCAK-- 752  
OY 269 VALP-----VRKSEPLQNVVDHMANHLGVSPKILLTGESELSPTAT--PSTKLGLVA 320  
DB 753 -KEPKDSCVTSPSEPVNVNPPSANPINIA-----IPVVAQLPTIYALADQ 798  
OY 321 DLIDCVVLASSSEATFSTQELRLVROGKEHQML-----EISLSP--DSPKLVLW- 368  
DB 799 NSVPC-----LRALANKOTILLIPOVAITYTSATVSPAMQEPKVIQIP 841



QY 369 -SHYEAMGLSGHKLSE---FPDGTGLSGKELPADLGLSGDLIE 409  
 DB 842 NGNODERODTSSGCVSEDDNDSDCTPPKKTKRAENGMYACDLCD 887

## RESULT 9

PSC\_DROME STANDARD; PRT; 1603 AA.  
 ID PSC\_DROME  
 AC P35820;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE POSTERIOR SEX COMBS PROTEIN.  
 GN PSC.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92018190; PubMed=1833647;  
 RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;  
 RT "Drosophila genes Posterior Sex Combs and Suppressor two of zeste  
 RT encode proteins with homology to the murine bmi-1 oncogene."  
 RL Nature 353:351-353(1991).  
 CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN  
 CC EXPRESSION PATTERNS OF THE HOMEOBOX SELECTOR GENES OF THE  
 CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR  
 CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLY).  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
 CC -----  
 CC DR EMBL; X59275; CAA1965.1; -  
 CC DR PIR; S17983; S17983.  
 CC DR FLYbase; FBgn0005624; Psc.  
 CC DR InterPro; IPR001841; Znf.ring.  
 CC DR Pfam; PF00097; zf-C3HC4; 1.  
 CC DR SMART; SM00184; RING; 1.  
 CC DR PROSITE; PS00518; ZINC\_FINGER\_C3HC4; 1.  
 CC KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.  
 CC FT DOMAIN 47 53  
 CC FT POLY-THR.  
 CC FT DOMAIN 83 88  
 CC FT POLY-THR.  
 CC FT DOMAIN 91 98  
 CC FT POLY-THR.  
 CC FT DOMAIN 145 152  
 CC FT POLY-THR.  
 CC FT DOMAIN 184 202  
 CC FT POLY-SER.  
 CC FT ZN\_FING 265 303  
 CC FT RING-TYPE.  
 CC FT DOMAIN 642 651  
 CC FT POLY-SER.  
 CC FT DOMAIN 1066 1069  
 CC FT POLY-GLY.  
 CC FT DOMAIN 1185 1189  
 CC FT POLY-PRO.  
 CC FT DOMAIN 1214 1217  
 CC FT POLY-PRO.  
 CC FT DOMAIN 1391 1396  
 CC FT POLY-PRO.  
 CC FT DOMAIN 1458 1461  
 CC FT POLY-ALA.  
 CC FT DOMAIN 1517 1520  
 CC FT POLY-GLY.  
 CC SQ SEQUENCE 1603 AA; 169999 MW; 77024F409736473 CRC64;

Query Match 5.7%; Score 119; DB 1; Length 1603;

Best Local Similarity 18.9%; Pred. No. 8.2; Mismatches 123; Indels 114; Gaps 13;

QY 45 VDLVSDSEVLEVADPEVVARLPAPAKPRDSDS---DSEGAEGPAGAPRTLVRRR 101  
 DB 598 IDLSKONSVTIIDMSPEPEREIVK---PLKPEKSRSKKKDKGSPKSSSSSSSSSSGGER 654

QY 102 RRRLDPGEAPVVPVYSGKVOSS---LNLIP-----DNSSLKLCPSPP----- 142  
 DB 655 KRKSPSLRPLPLITRTERIMSPSGVSTLSPRYTSGAFSDPSPKSEFLKSPALKPIKAYE 714  
 QY 143 EDEADLTNNGSSSPSEDALPSSGSPWRRKKLRKCKEKEKKKEEPPDDDISPLPSSRNKS 202  
 DB 715 SPRTLNINRAITPPSPSVQOASAP-----KSKGNLDDSIILMPPCMPKRSIASSKRS 769  
 QY 203 RKHTPALQKLRVKNRKLODRSLSPKQHOAPALQSDDEVYL-----VEG 248  
 DB 770 KEPPKAVSK-----KQKLSPLPVDFKIRLPVTNGSSGTASPKEIK 812  
 QY 249 PVLQSSR---LFTLIRCRADVLRLP-----YRMSRP-----LQNVVDHMA- 287  
 DB 813 PLMPRPAPKPPMLAPRLQSPAGAPAPPSPIHHHAGVQMSAPGNRTPIAKRYOPIILKASR 872  
 QY 288 -NLGVSPNRILLF-----GESLSPAPAPST 314  
 DB 873 PNPFANIPNDVNRLLKDACGTEIKSIGGSVENNSNSAQKPHLYGPKGETKMGPPALPAT 931

## RESULT 10

MSL1\_DROME STANDARD; PRT; 1039 AA.

ID MSL1\_DROME  
 AC P50535;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE MALE-SPECIFIC LETHAL-1 PROTEIN.  
 GN MSL-1.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN (1)  
 RP SEQUENCE OF 85-1039 FROM N.A.  
 RX STRAIN-CANTON-S;  
 RC MEDLINE=93314941; PubMed=8325488;  
 RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,  
 RA Luccchesi J.C.;  
 RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster  
 RT encodes a novel protein that associates with the X chromosome in  
 RT males."  
 RL Genetics 134:545-557(1993).  
 RN [2]  
 RP REVISIONS, SEQUENCE FROM N.A.  
 RX MEDLINE=95300219; PubMed=7781064;  
 RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,  
 RA Kuroda M.I.;  
 RT "Expression of msl-2 causes assembly of dosage compensation  
 RT regulators on the X chromosomes and female lethality in Drosophila."  
 RL Cell 81:867-877(1995).  
 CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING  
 CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME  
 CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND  
 CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL  
 CC PROTEINS REQUIRES ALL THE OTHER MSLS FOR WILD-TYPE X-CHROMOSOME  
 CC BINDING.  
 CC -1- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS  
 CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES  
 CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL  
 CC SITES IN MALES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
 CC -----

DR	EMBL: LA2514; AAA6918.1 -	
DR	FLYBase: FBgn0005617; ms1.1.	
FT	Nuclear protein.	
FT	CONFLICT 188	193
FT	CONFLICT 492	492
FT	CONFLICT 670	670
FO	SEQUENCE 1039 AA; 117412 MW; 4759E6B5EEF5F14 CRC64;	

Query Match	5.6%;	Score 116.5;	DB 1;	Length 1039;
Best Local Similarity	22.2%;	Pred. No. 6.9;		
Matches	82;	Conservative 55;	Mismatches 141;	Indels 91;
				Gaps 19;

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

CC BE INTRODUCED IN POSITION 944TO PRODUCE THIS ORF.  
CC  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL, X62629; CAA44496.1; ALT_FRAME.
DR	PIR: A40580; A40580.
DR	FlyBase; FBgn0002542; 1d5.
DR	InterPro: IPR001410; DEAD.
DR	InterPro: IPR001650; Helicase_C.
DR	InterPro: IPR000350; SNEF_N.
DR	Pfam; PF00271; helicase_C; 1.
DR	Pfam; PF00176; SNEF_N; 1.
DR	SMART; SM00487; DEXDC; 1.
DR	SMART; SM00490; HELIC; 1.
KW	Nuclear protein; Helicase; ATP-binding.
FT	NP_BIND 465 472
FT	SITE 603 606
SE	SEQUENCE 1061 AA; 118189 MW; E82AA66254342B80 CRC64

Query Match	5.6%;	Score 116.5;	DB 1;	Length 1061;
Best Local Similarity	22.2%;	Pred. No. 7;		
Matches 92;	Conservative 50;	Mismatches 144;	Indels 129;	Gaps 20;

[illegible]

RESULT	12		
TCF8_MESAU			
ID	TCF8_MESAU	STANDARD:	PRT: 1043 AA.
AC	060542:		
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	TRANSCRIPTION FACTOR 8 (ZINC FINGER PROTEIN BZP).		
GN	TCF8 OR BZP.		
OS	Mesocricetus auratus (Golden hamster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;		
OC	Mesocricetus.		
OX	NCBI_TaxID=10036;		



```
FT REPEAT 608 619 2-4.
FT REPEAT 623 634 2-5.
FT REPEAT 636 647 2-5.
FT REPEAT 650 661 2-7.
FT REPEAT 683 694 2-8.
FT REPEAT 717 728 2-9.
SQ SEQUENCE 872 AA; 97305 MW; C60F5BE8808ED1D31 CRC64;

Query Match 5.5%; Score 115.5; DB 1; Length 872;
Best Local Similarity 24.5%; Pred. No. 6.4;
Matches 87; Conservative 43; Mismatches 128; Indels 97; Gaps 17;

QY 47 LVSDDEEVLEAVDEVPVAPARLPAPAKP-EQDSDSSEGAEGPAGAPRTLVRRRRRRL 105
D 215 LSSEHHEVEVEVEDDD-----SSAKTGQKVDPSFASILLGKTRKRVRRIRK 267
QY 106 LDPGAPVVPYVSGVQSS-----LNLIPNSSLKLCPEPPDEADLTSSGSSP 155
D 268 -----NFKSKRVRESEHTTFQDPNPLNQSSNNSEARQDDADE-DQDSNNDSP 317
QY 156 SEDALPGSPWRKRLRCKEKEKMEFPQDISPLPPQSSRKSRKHTFALQKLEEV 215
D 318 L-DEFLP-----MDQLKRLYKRRKN-----SGLVSSLPSEQ-----QETREE 354
QY 216 NKRLQDLRSCUSP-KOHOSPALQSTDEEVVVEGFPVLQOSSRLFTLKTGRADVLRLPVR 274
D 355 KKVLEDMKDSLSHFQK-----IQYVDSASLPISSVFLQNGNTLPTSNVNTTVPQQLPL- 408
QY 275 MSEPLONVVDHANHL-----GVSPLRILLPGESELSPAT-----PST 314
D 409 --EPKLPATGSAHNLVBEYNQHPNSGALQTLQPLKPTATGSANYLKMRSHQPOPS 466
QY 315 LKLGVAADIIDCVLASSEATETS---QELRLRYQKEKHOMLE-LSLSPDSP 364
D 467 IK-----PSSITPEYVTVNGSGIQLKPTATGSANYLKMRSHQPSVNPV 509

RESULT 14
TCOF_HUMAN STANDARD; PRT; 1411 AA.
AC 013428; 099408;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRACHELE PROTEIN (TREACHER COLLINS SYNDROME PROTEIN).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA Bonner C.A., Kopitvnikar K., Wasmuth J.J.,
RA "Positional cloning of a gene involved in the pathogenesis of
RA Treacher Collins syndrome. The Treacher Collins syndrome
RA Collaborative Group."
RL Nat. Genet. 12:130-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nuclear phosphoprotein that exhibits
RT mutations in Treacher Collins syndrome throughout its coding
RT region."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [3]
RP VARIANTS L-439; V-810; V-1313 & G-1355, AND VARIANT TCS R-53.
RX MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
```

```
RT "The mutational spectrum in Treacher Collins syndrome reveals a
RT predominance of mutations that create a premature-termination
RT codon."
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS
CC SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF
CC CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000
CC LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY
CC SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,
CC AREAS OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE
CC EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)
CC LATERAL DOWNGRADE SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH
CC COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND
CC ZYGOMATIC COMPLEX; (4) CLEFT PALATE.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40847; AAC50903.1; -;
DR EMBL; U76366; AAC51181.1; -;
DR EMBL; U84664; AAC51185.1; -;
DR EMBL; U84640; AAC51185.1; JOINED.
DR EMBL; U84641; AAC51185.1; JOINED.
DR EMBL; U84642; AAC51185.1; JOINED.
DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.
DR EMBL; U84655; AAC51185.1; JOINED.
DR EMBL; U84656; AAC51185.1; JOINED.
DR EMBL; U84657; AAC51185.1; JOINED.
DR EMBL; U84658; AAC51185.1; JOINED.
DR EMBL; U84659; AAC51185.1; JOINED.
DR EMBL; U84660; AAC51185.1; JOINED.
DR EMBL; U84661; AAC51185.1; JOINED.
DR EMBL; U84662; AAC51185.1; JOINED.
DR EMBL; U84663; AAC51185.1; JOINED.
DR MIM; 154500; -.
KW Disease mutation; Polymorphism.
FT DOMAIN 89 97
FT DOMAIN 204 207
FT DOMAIN 616 619
FT DOMAIN 919 924
FT DOMAIN 1285 1289
FT DOMAIN 1375 1386
FT DOMAIN 1398 1405
FT VARIANT 53 53 W -> R (IN TCS).
FT VARIANT 439 439 /FTId=VAR_005630.
FT VARIANT 810 810 P -> L.
FT VARIANT 810 810 /FTId=VAR_005631.
FT VARIANT 1313 1313 A -> V.
FT VARIANT 1313 1313 /FTId=VAR_005632.
FT VARIANT 1355 1355 A -> V.
FT VARIANT 1355 1355 /FTId=VAR_005633.
FT VARIANT 1355 1355 D -> G.
SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 5.5%; Score 114.5; DB 1; Length 1411;
```

Best Local Similarity 21.4%; Pred. No. 13;  
Matches 84; Conservative 50; Mismatches 172; Indels 87; Gaps 16;  
QY 2 AEPLRGRGPRSG-----RGARARGARGRCRRARQSPARLIPDVIVLVSDEEVL 56  
DB 715 ANPAARASAKGTISAPGVVTAQAQORSPSKVPPV-----NPNSTV 762  
QY 57 EVADPEVEP-VARLAPAPAK-----PEQDS-----DSDEGAAGPAGPRTIVRRRRRL 105  
DB 763 LANGPASVSVGKAVATAQAQGTPEDESGSSSEESDEEATLQAQKPSGKTHQIRAA 822  
QY 106 LDP-----GEAPVVPVYSGKVOSSLNLPDSSLLKLCPSPEDEADITNSG----- 152  
DB 823 LAPAKESPCKGAPPTPGKTPGSAQAQKODSGSSSEESDEGEAPAAVTSQVTKPPL 882  
QY 153 -SSPEEDALPSGSPWR-----KKLRKCKEKKKMEFEFDODISLPQ---SSRN 200  
DB 883 IFVDPNRSPGAPATPAQAQASTPRKARSESTANSSSESEDEVDIPATQCLTPTGIRT 942  
QY 201 K-----SRKHTALQKLEVNKRLODLSCLSPKOHSPALQSTDEVLVEGP 249  
DB 943 NYVTMTAPRIAPKASMGASSKSSRI SD-----GKQSGPATQ-----VSKNP 990  
QY 250 V-LPOSSRLFTLKIRC-RADLVRLPYRMSEPLQNVVDMANHLGVSPNRIILFGESELS 307  
DB 991 ASLPLTQA--ALKVLQAQKASEAQPPVARTOPSSGVDSAVGTLPATSPQ-----STSVQ 1041  
QY 308 PIATPSTLKGVAIIDICVVLASSSEATETSOE 340  
DB 1042 AKGTNKLKPKLPEVOATKAPESSDSDSDSD 1074  
RESULT 15  
DYNA\_RAT  
ID DYNA\_RAT STANDARD: PRT: 1280 AA.  
AC P28023;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE DYNACTIN 1 (150 KDA DYNEIN-ASSOCIATED POLYPEPTIDE) (DP-150) (DAP-150)  
DE (P150-GUED).  
GN DCTNI.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RX MEDLINE=91260877; PubMed=1828535;  
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitz N.G.,  
RA Pfister K.K., Vallee R.B.,  
RT "Homology of a 150k cytoplasmic dynein-associated polypeptide with  
RT the Drosophila gene Glued".  
RL Nature 351:579-580(1991).  
RN [2]  
RP REVISIONS.  
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitz N.G.,  
RA Pfister K.K., Vallee R.B.,  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE  
CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-  
CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL  
CC TRANSPORT OF VESICLES AND ORGANELLES.  
CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.  
CC P150(GUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC  
CC DYNEIN.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
CC -!- SIMILARITY: STRONG, TO OTHER SPECIES DYNACTIN 150 KDA SUBUNIT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X6160; CAA44091.1; -  
DR PIR; S16129; S16129.  
DR InterPro; IPR000938; CAP-Gly.  
DR Pfam; PF01302; CAP\_GLY; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.  
FT DOMAIN 48 90  
FT DOMAIN 157 184  
FT DOMAIN 214 513  
FT DOMAIN 942 1048  
FT DOMAIN 1184 1213  
SQ SEQUENCE 1280 AA; 141929 MW; C9348CF129FAFF5C CRC64;

Query Match 5.4%; Score 113.5; DB 1; Length 1280;  
Best Local Similarity 22.6%; Pred. No. 13;  
Matches 87; Conservative 61; Mismatches 144; Indels 93; Gaps 17;

QY 71 APAKPEQSDSDSEGAEGPAGAPRTIVRRRRRLDPGEAPV-----PYVS 118  
DB 105 SPETP---DSSASKILKREGADAAKT-----SKRLGKPKKAPAKRTTTRRRPTPAST 158  
QY 119 GKVOSSLNLPDSS-SLKLKCPSEPEDEAD-----LTNCGSSPEEDALPSGS 165  
DB 159 GVGAPSSSLGPGSGSASAGELSSSESTPATQPLAAPLIPTPALITSGAAP---PLPSPS 214  
QY 166 PMRKKLRKCKEKKKMEEPDODISLPPOSSRNKR-----KHTDALQKLEVNKRQ 220  
DB 215 KEEEGLRDVRDLEKLE-----TLRLKRSDEKAKLEKHKIOLQGVQEWKSKMQ 266  
QY 221 ---DIRSCLSPKOHSPALQSTD-----DEVVLVEGVLPOSSRLFTLKIRCRADLV 269  
DB 267 EQGADQLRLKLEKKEKKEKLEKERYMEEMADPADIAEATLDKE-----NAEERAESE 320  
QY 270 RLPEVRMSEPLQNVVDMANHLGVSPNRIILFGESELSPTAPPS---TLKLGVAIIDIC 325  
DB 321 QQEV---EALKERVDELTTDLLELKAEI-----EEKGSDGAASSYQKOLEQGNARLKDA 372  
QY 326 VYVASSSEATFNSQELRLVQGEKKEKHOMLISLPSPLKVLMSHVEAMGLSGHLSFF 385  
DB 373 LVAMRDLSSSEKQEHVKLQKLEKKNOLEVVNRQORERLOEELSQAESTI----- 422  
QY 386 FDGTRKLGKELPADLGLSGDLIEV 410  
DB 423 -DELK---EQVDAALGAE--EWYEM 441

Search completed: April 22, 2002, 10:09:19  
Job time: 184 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:07:05 ; Search time 26.46 seconds  
(without alignments)  
2277.559 Million cell updates/sec

Title: US-09-617-923-2  
Perfect score: 2099  
Sequence: 1 MAEPLRGGRGPRSRGRRGARR.....GKELPADLGESGLIEYWG 412

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SPREMBL\_17:\*
  - 2: sp\_archaea:\*
  - 3: sp\_bacteria:\*
  - 4: sp\_fungi:\*
  - 5: sp\_human:\*
  - 6: sp\_invertebrate:\*
  - 7: sp\_mhc:\*
  - 8: sp\_mammal:\*
  - 9: sp\_organelle:\*
  - 10: sp\_phage:\*
  - 11: sp\_plant:\*
  - 12: sp\_rodent:\*
  - 13: sp\_virus:\*
  - 14: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	11	009130 mus musculu
2	1982	94.4	416	11	09CIV5
3	1382	65.8	408	6	09GIZ9
4	154.5	7.4	1217	4	060336
5	145.5	6.9	1072	11	035482
6	140	6.7	1634	5	09XW25
7	133	6.3	971	5	09XVS4
8	132.5	6.3	792	5	018866
9	131	6.2	664	4	092541
10	130	6.2	704	4	09H5F9
11	128.5	6.1	484	11	09CS98
12	128.5	6.1	635	10	09C363
13	128.5	6.1	712	5	061708
14	128.5	6.1	755	5	09Y4J5
15	128.5	6.1	931	13	091995
16	127.5	6.1	1167	5	09VNA7
17	127.5	6.1	1408	5	09W0C9
18	127	6.1	239	12	057114
19	127	6.1	1368	5	09V6J0

20	126	6.0	856	5	021186	021186 caenorhabd
21	125	6.0	390	5	09XUJ9	09XUJ9 caenorhabd
22	124.5	5.9	670	4	09BW08	09BW08 homo sapien
23	124.5	5.9	757	13	09YHD2	09YHD2 gallus gall
24	124.5	5.9	827	4	09BRK5	09BRK5 homo sapien
25	124.5	5.9	879	2	09FBJ2	09FBJ2 streptomyce
26	124.5	5.9	1340	2	0911H8	0911H8 streptomyce
27	124.5	5.9	1783	4	015038	015038 homo sapien
28	124.5	5.9	1791	4	060382	060382 homo sapien
29	124.5	5.9	2286	4	09URH8	09URH8 homo sapien
30	123.5	5.9	802	4	09UQ37	09UQ37 homo sapien
31	123.5	5.9	815	11	09ERQ2	09ERQ2 rattus norv
32	123.5	5.9	2752	4	09UQ35	09UQ35 homo sapien
33	122.5	5.8	1152	5	09VZ23	09VZ23 streptomyce
34	122	5.8	876	11	09PC22	09PC22 mus musculu
35	122	5.8	1082	5	001905	001905 caenorhabd
36	122	5.8	1601	5	09V619	09V619 drosophila
37	121.5	5.8	678	11	099L92	099L92 mus musculu
38	121.5	5.8	1810	5	09W363	09W363 drosophila
39	121	5.8	584	11	09D672	09D672 mus musculu
40	121	5.8	886	13	P70049	P70049 xenopus lae
41	121	5.8	1668	4	015026	015026 homo sapien
42	121	5.8	2971	4	0915L9	0915L9 homo sapien
43	120.5	5.8	6994	5	017343	017343 caenorhabd
44	120.5	5.7	1136	13	098UH3	098UH3 xenopus lae
45	120.5	5.7	1881	2	0917Q2	0917Q2 streptococc

ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	412 AA.
009130				
AC	009130			
DT	01-JUL-1997 (Tremblrel. 04, Created)			
DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE	NUCLEAR PROTEIN NIP45.			
GN	NEATC2IP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI	Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97102593; PubMed=8943202;			
RT	Hodge M., Chun H., Rengaraajan J., Alt A., Lieberman R., Glimcher L.,			
RT	"NF-AT-Driven Interleukin-4 transcription potentiated by NIP45.";			
RL	Science 274:1903-1905(1996).			
DR	EMBL: U76759; AAC52963.1; -			
DR	MGI: MGI:1329015; Nfatc2ip.			
DR	InterPro: IPR000626; Ubiquitin.			
DR	SMART: SM00213; UBO; 1.			
DR	PROSITE: PS50053; UBIDUTIN_2; 1.			
KW	Nuclear protein.			
SO	SEQUENCE 412 AA; 45121 MW; DD58F5C7055C186 CRC64;			

Query Match 100.0%; Score 2099; DB 11; Length 412;  
Best Local Similarity 100.0%; Pred. No. 4e-143;  
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEPLRGGRGPRSRGRRGARRAGRCPRAROSPALLIDYLVLDVVSDEVELEVAD	60
DB	1	MAEPLRGGRGPRSRGRRGARRAGRCPRAROSPALLIDYLVLDVVSDEVELEVAD	60
QY	61	PVEVPVAPRLPAPAKPPEDSDSEGAEPAGAPRTLVRRRRRLDPGAPVVPVYSGK	120
DB	61	PVEVPVAPRLPAPAKPPEDSDSDSEGAEPAGAPRTLVRRRRRLDPGAPVVPVYSGK	120
QY	121	VQSSLNLIIPDNSSLLKLCPEPEDEADLTNSGSSPSSEDDALPGSGPWRRKLRRKCEKEK	180
DB	121	VQSSLNLIIPDNSSLLKLCPEPEDEADLTNSGSSPSSEDDALPGSGPWRRKLRRKCEKEK	180

Db 121 VOSSLNLIDNSSLLKLCSEPEDEADLTNCGSSPEDEADLPSGSPWKKLRKCKEERK 180  
QY 181 KNEEFPDODISLPPOSSRNKSRKHTALQKLEVNKRLODLRSCLSPKOHOSPALQSTND 240  
Db 181 KNEEFPDODISLPPOSSRNKSRKHTALQKLEVNKRLODLRSCLSPKOHOSPALQSTND 240  
QY 241 DEVLVEGVLPQSSRLFLTKRCRADLVRLPYRMSEPLQNVDMANHLGVSPNRILL 300  
Db 241 DEVLVEGVLPQSSRLFLTKRCRADLVRLPYRMSEPLQNVDMANHLGVSPNRILL 300  
QY 301 FGESELPATPSTLKLGVADIIDCVLASSSEATETSOELRLRQGEKHOEITSLSP 360  
Db 301 FGESELPATPSTLKLGVADIIDCVLASSSEATETSOELRLRQGEKHOEITSLSP 360  
QY 361 DSPPLKVMHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 412  
Db 361 DSPPLKVMHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 412

RESULT 2

Q9CVY5 PRELIMINARY: PRT: 416 AA.  
AC 09CVY5: 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DE NUCLEAR FACTOR OF ACTIVATED T-CELLS, CYTOPLASMIC 2 INTERACTING PROTEIN (FRAGMENT).  
GN NFATC2IP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleschmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo T., Nakai I., Pesole G., Quackenbush J., Schiml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombaerts P., Nori P., Rong B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wushaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayschizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL: AK005947; BAB24331.1; -.  
DR MGD: MGI:1329015; Nfatc2ip.  
DR InterPro: IPR000626; Ubiquitin.  
DR SMART: SM00213; UBO; 1.  
FT NON\_TER 1  
SQ SEQUENCE 416 AA; 45551 MW; E4B46F65CC571AF5 CRC64;

Query Match 94.4%; Score 1982; DB 11; Length 416;  
Best Local Similarity 97.0%; Pred. No. 1e-134;  
Matches 394; Conservative 1; Mismatches 9; Indels 2; Gaps 1;  
QY 9 GPSRSGRGARRA--RGARGRCPRAROSPARIIPDTVLDVLDSDSEVLEVPVPEVY 66  
Db 11 GTMSEVPKPRPRSRRTGANGRCPRAROSPARIIPDTVLDVLDSDSEVLEVPVPEVY 70

QY 67 ARLPAPAKPEODSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAVPVVYSGKVOSSLN 126  
Db 71 ARLPAPAKPEODSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAVPVVYSGKVOSSLN 130  
QY 127 LIPDSSLKLCPSPEDEADLTNCGSSPEDEADLPSGSPWKKLRKCKEERKNEEFP 186  
Db 131 LIPDSSLKLCPSPEDEADLTNCGSSPEDEADLPSGSPWKKLRKCKEERKNEEFP 190  
QY 187 DODISLPPOSSRNKSRKHTALQKLEVNKRLODLRSCLSPKOHOSPALQSTNDDEVVLY 246  
Db 191 DODISLPPOSSRNKSRKHTALQKLEVNKRLODLRSCLSPKOHOSPALQSTNDDEVVLY 250  
QY 247 EGVLPQSSRLFLTKRCRADLVRLPYRMSEPLQNVDMANHLGVSPNRILLFGESEL 306  
Db 251 EGVLPQSSRLFLTKRCRADLVRLPYRMSEPLQNVDMANHLGVSPNRILLFGESEL 310  
QY 307 SPAPSPSTLKLGVADIIDCVLASSSEATETSOELRLRQGEKHOEITSLSPSPKLV 366  
Db 311 SPAPSPSTLKLGVADIIDCVLASSSEATETSOELRLRQGEKHOEITSLSPSPKLV 370  
QY 367 LMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 412  
Db 371 LMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 416

RESULT 3

Q9GLZ9 PRELIMINARY: PRT: 408 AA.  
AC 09GLZ9: 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DE HYPOTHETICAL 44.6 KDA PROTEIN.  
OS Macaca fascicularis (Craab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRN PARIETAL LOBE;  
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;  
RT Isolation of full-length cDNA clones from macaque brain cDNA libraries.  
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AB050511; BAB17279.1; -.  
DR InterPro: IPR000626; Ubiquitin.  
DR SMART: SM00213; UBO; 2.  
DR PROSITE: PS50053; Ubiquitin.  
KW Hypothetical protein.  
SQ SEQUENCE 408 AA; 44580 MW; 5E0DD0D066AC2F24 CRC64;

Query Match 65.8%; Score 1382; DB 6; Length 408;  
Best Local Similarity 70.1%; Pred. No. 1.3e-91;  
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

QY 1 MAEPLNGRGPRSGRGARRANGRCPRAROSPARIIPDTVLDVLDSDSEVLEVY 58  
Db 1 MAEPLNGRGPRSGRGARRANGRCPRAROSPARIIPDTVLDVLDSDSEVLEVY 52  
QY 59 ---ADPVEVPYARLPAPAKPEODSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAVPV 114  
Db 53 ARCAADVEVAPSEPPGYPVARSRDDSDSDSEGADARAGPRPREVRRRRRLVDPGEAPLV 112  
QY 115 PYSGKVOSSLNIPDSSLKLCPSPEDEADLTNCGSSPEDEADLPSGSPWKKLRK 174  
Db 113 PYSGKVOSSLNIPDSSLKLCPSPEDEADLTNCGSSPEDEADLPSGSPWKKLRK 171  
QY 175 CEKEEKKNEEFPDODISLPPOSSRNKSRKHTALQKLEVNKRLODLRSCLSPKOHOSP 234  
Db 172 -DKEEKKTEIILDNSSLSPSPSPRTKSRKHRAKLKLEVNKRLODLRSCLSPKOHOSP 230



```

OY 235 ALSTDDDEVLVVEGVGLPOSSRLFTFKICRGADIVRLPYRMSPEFLQWVNDHMANHLGVSP 234
DB 231 RQGGDEDEVLVVEGVGLPEFRPLPLKICRGADIVRLPYRMSPEFLQWVNDHMANHLGVSP 230
OY 295 NRILLFGESESLPTAPFPSTLKLGVADIIDCVVLASSSEATEETSOELRLRVQGEKHQML 354
DB 291 SRILLFGESESLPTAPFPSTLKLGVADIIDCVVLASSPEATEETSRQLQLRVQGEKHQTL 350
OY 355 ELSLSPDSPLVLMHSHEAMGLSGHKLSFFDGTKLSGKELPADLGLSGDLIEWMG 412
DB 351 EVSLRSRSPKLTMLMSHYEAMGLSGRKLSPFFDGTKLSGRELPAIDLGMSGDLEIYEWG 408

RESULT 4
060336 PRELIMINARY; PRT; 1217 AA.
AC 060336;
DT 01-AUG-1998 (TREMBLrel, 07, Created)
DT 01-AUG-1998 (TREMBLrel, 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE KIAA0596 PROTEIN (FRAGMENT).
GN KIAA0596.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011168; BAA25522.1; -.
DR InterPro: IPR002114; PTS_HPr_ser.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 1217 AA; 131097 MW; 0BC4EAC6722BEF5 CRC64;

```

```

Query Match          7.48; Score 154.5; Ds 4; Length 1217;
Best Local Similarity 23.58; Pred. No. 0.0085;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps

Qy 6 RGRPSRSGKGGARRAGRCRCRARSARLIPDTVLVDIVSDSEETLEVAADPEYV 65
   | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 455 RORGGKQOGPSSPRAG-----PNRIQATSMLSPGA---LSDSDKKEEDCTEETELP 506
                                     | : | : | : | : | : | : | : | : | : |
Qy 66 -----VALPAPAKPEODSDSDSEGAEE-----PGAPAPRLTVRRRRRL 106
                                     | : | : | : | : | : | : | : | : | : |
Db 507 ALPIYAKSTKALASVPSPALPRSLTSMKMSRAQESVGLFDPAAPANPGP---RRGRRW 563
                                     | : | : | : | : | : | : | : | : | : |
Qy 107 DPGELPVVYVYSGVQSSLANLIPNSSLKLCPS--EP-EDEADLTNCGSSPSPEDALPS 163
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 564 QPG-----VELSVRSMDLRLRLTSLAPSLQPSODSLAIIPSGRKHGOEALFT 612
                                     | : | : | : | : | : | : | : | : | : |
Qy 164 G-----SPWKKLKRKC-----EKKEE-----KKMEEPPODISPLRPPS---SRNK 201
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 613 SLTSONKEPPPPQASQPCSTPHIIRLLSOEBGVFAODLEAPFEDGIVVPEPSDNPMPOT 672
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 202 SRKHTEALQ-----LREVNKRLOD-----LRSLCSKHQSPALQSTDEDEVYL 245
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db      673  SEIQQAARATTLGRVYPGSSSEKHPDCAKSVDISSCLSPSEH--PIEDSESTPLS 730
OY      246  VEGPVLPOSSRLFTLKIRCADLVRPRMSEPLQNVYDHANHLGVSPRILLFGESE 305
Db      731  VDG-----ISDLEEPAGDEEEEBEGGMP-----YGLQE 762
OY      306  LSPRTPT-----STLKLGVADIIIDCVLLASSEAEIETSOELRLRQGEKKNOMLE 355
Db      763  GSP-OTPRQEOFLKHQETETLASCAGAPAP-VQYPERESRSISSRFLQVOTPRREP-- 818
OY      356  ISLSPDSPKLKVLMSHYEAMGLSHK 381
Db      819  ---SPSSSTALMSRPAPVQASSEQ 841

```

```

RESULT      5
035482      PRELIMINARY;      PRT;      1072 AA.
AC      035482;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE      HIGH MOLECULAR-WEIGHT NEUROFILAMENT.
GN      NF-H.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91038277; PubMed=2230956;
      Chin S.S., Liem R.K.;
RT      "Transfected rat high-molecular-weight neurofilament (NF-H)
      cossmbles with vimentin in a predominantly nonphosphorylated form."
RL      J. Neurosci. 10:3714-3726(1990).
      [2]
RP      SEQUENCE FROM N.A.
RA      Chin S.S.M., Liem R.K.H.;
      Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC      EMBL; AF031879; AAB87068.1; -.
DR      InterPro: IPR001664; IF.
DR      Pfam: PF00038; filament.1.
DR      PROSITE: PS00226; IF, 1.
KW      Colled coil; Intermediate filament.
SQ      SEQUENCE 1072 AA: 115348 MW: 89A146D57A4D78B CRC64;

```

	Query Match	6.9%	Score 145.5:	DB 11;	Length 1072;	
	Best Local Similarity	22.1%;	Pred. No. 0.03;			
	Matches	95;	Conservative	63;	Mismatches 166;	Indels 105; Gaps 20.
QY	4 PLRGHPRSGRGGARRARGARNGRCRRARQSPARL-IPDVLVDLVSDSEEV-----	55				
Db	613 PVEAKSPAEPKSPAVKSPG--EAKSPAEPKSPAEVKSPTATVKSPEAKSPAEPKSPVTWK	671				
QY	56 --LEVADVEVYVALPAPAKPEQDSDDSEGAAGCAPARTIVRRRRRLLDGEAAV	113				
Db	672 SPAEPKSPVEV----KSAVSYSPESEAKSPAGAASPAEKASPVAKSPAEPKSPAEPK-	725				
QY	114 VPVYSGKVSSINLLIPDNSSLILCLPPSEPEDADLTN-----SGSSPSDDA---LP	162				
Db	726 -PPAAKSPAEPKSPAEPK-----PAEKSPAEPKSPVEVKSPEKASPVEKAQSLAE	779				
QY	163 SGSPARKLRKCKEEKKMEFPPODISPLPQP-SRKRSR-----KHTEALQKLREYVK	217				
Db	780 AKSPEKASPVKEELIKPAEVKSPKASPMEEKASKPKATLDVKSPEATTPAKEAK	839				
QY	218 RLQDLRSLSPKQHOSPALQSTDEDEVLEVGVLSQSSRLFLTKIRCRADLYRLP---VR	274				
Db	840 RPADI---SPEOVKSPAEPKES-----PEKETETREKVAAPKKEEVKSPEVEPK	886				
QY	275 MSEPIONTVVHDHANLIGSPNRILLFGESLSPATATPETTKLGVAADITDCVLIASSSPA	334				

```
Db 887 ANEP-----PKRV-----SEKTP-ATPKTE-----KESKDDA 915
Oy 335 TTSOELRLVQGEKHKOMLEISLSPDPLKYLMSHYEAMGLSGHLSFFDGTLSGK 394
Db 916 PKEAO-----KPKAEKEPILTE--KPKDSGCEAKKEAKKKAAP-----E 956
Oy 395 ELPADIGLE 403
Db 957 EPPAKLGVK 965

RESULT 6
O9XW25 PRELIMINARY; PRT; 1634 AA.
AC O9XW25;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Y18D10A.1 PROTEIN.
GN Y18D10A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.;
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latrelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Shownkeen R.,
RA Smaidon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; AL034393; CAA22308.1; -.
DR InterPro; IPR000637; AT_hook.
DR PROSITE; PS00354; HMGL.Y; UNKNOWN_1.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 1634 AA; 179058 MW; DC20372F3AAF48D0 CRC64;
```

```
Db 643 RLDDAKTTATYIHPPPLTRKMERKAPPAVTSSKKEPKNAGSDSSINEEHEDE 702
Oy 243 VLVAGPV--LPQSRRLFTLIKIRCRADLVRLVPRASEPLQWVDMAHNLGVSPRILL 300
Db 703 TMLEEQILDLPOQPSQOEPRISGSEL-----LDGEFDSESHSGTVPS----- 747
Oy 301 RGESELSPTAPSTIKLGVADIICVILASSEATETSQE 340
Db 748 --APELTRKNPAPPV-----PEASEASAE 768

RESULT 7
O9XVS4 PRELIMINARY; PRT; 971 AA.
AC O9XVS4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE C25A1.10 PROTEIN.
GN C25A1.10
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latrelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Shownkeen R.,
RA Smaidon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z81038; CAB02755.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 971 AA; 101096 MW; 832936115B58709B CRC64;
```

Query Match 6.3%; Score 133; DB 5; Length 971;  
Best Local Similarity 23.3%; Pred. No. 0.23;  
Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

```
Oy 29 PARQSPALLIPDYVLVDS---DSDEVLEVADPEVPVRLAPAKPEDSDSDSG 85
Db 231 PAKPTPAKPTPKPVYKKAESSSDSDDEKKPVAKP--APAKATPKPAKKADSSSDSD 288
Oy 86 ----AAEGPAGAPRTLVRRRRLD-----PGEAPVVPYSGKVOSSLNLI 128
Db 289 DEAPPAKTPAKAPKPVAKKAESSSDSDDEKKPAKPTPAKATPKPVAKKAESSSDSD 348
Oy 129 PNNSLLKLKPS--PEDEADLTNCGSPSEDALPSGSPW--RKLRLKCKEKKMEEF 185
Db 349 DEKKPVAKPAKAPKAPKPVAKKAESSSDSDDEKKPAKPTPAKATPKPVAKKAESSSDS 408
Oy 186 PPODISPLPQPSRNKS-----KHTFALQKLRVKNRLDLSNC 225
Db 409 SDDEKKPVAKPTSAKTPPAKAKKADSSSDDEAPPAKTPAKAPKAPKAKAE---SS 465
Oy 226 LSPKHQSPALOSTDEV 243
```



RESULT	ID	Q9H5F9	PRELIMINARY;	PRT;	704 AA.
Db	387	DQRFVRLFEVSNQEFTESEFMK-WKEAFMSAQM0LPTL---	DEINKKEL	431	
RESULT	10				
Q9H5F9	AC	Q9H5F9	PRELIMINARY;	PRT;	704 AA.
Db	01-MAR-2001	(TREMBLrel. 16, Created)			
Db	01-MAR-2001	(TREMBLrel. 16, Last sequence update)			
Db	01-JUN-2001	(TREMBLrel. 17, last annotation update)			
De	CDNA:	FLJ23471 FIS, CLONE HS111969.			
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
NCBI_TaxID=9606;					
SEQUENCE FROM N.A.					
RA	TISSUE=HUMAN SMALL INTESTINE;				
RA	Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,				
RA	Okltani R., Ota T., Suzuki Y., Ohsayashi M., Nishi T., Shibahara T.,				
RA	Yanaka T., Nakamura Y., Isozaki T., Sugano S.;				
RT	"NEBO human cDNA sequencing project."				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AK027124; BAB15667.1; -				
DR	InterPro: IPR002965: P-rich extensin.				
DR	PRINTS: PRO1217; PRICHEXTENS.				
DR	SEQUENCE 704 AA: 75768 MW; 856F94EAA2C1F8C6 CRC64;				
Query Match		6.2%; Score 130; DB 4; Length 704;			
Best Local Similarity		22.6%; Pred. NC. 0.25; Indels 106; Gaps 20;			
Matches	100; Conservative	63; Mismatches 173; Indels 106; Gaps 20;			
QY	10	PSRSGRGARARAGARGRCRPARQSPARLIPTVIVDLYSDSEYLEVADPV-----	62		
Db	286	PTEAPQASPLAKPLQSSSPVTLGLPSRNEPAPL---STSTQSASLPPAGRNLAE	341		
QY	63	-----EVYARLPAPAKPRQSDSDSEGAEPAG-----AP	94		
Db	342	SSGVGRVAGSGRPKEAPAKGKSTLTITQDMSTSLQEGEDGPGWRANLKEVDRKSPAE	401		
QY	95	RLTVRRRRRLDP--GEAP--VVPYSGKVOSSLIL-I-PDMS-----SLTKCPSE	141		
Db	402	RLTKPEPALPEPRAGEPRKRVSGSFASVHITLTPVRDTPRPASGSPSPARSPSP	461		
QY	142	PRDEADLTNGSSPSSEDDALPGSPWRKRLRKCKEKK--MEEPDQDISP-----LP	194		
Db	462	PR-RRRLANPAISLDVCDMLRPPEPQGEARVQSKWEKKPHLQGGKGRPLSPANPALP	520		
QY	195	QPSRRKSKKHTBALQKLEVNKRILQDLRSCLSPKQHOSPALQST-----DD--EVILV	246		
Db	521	GETVISPVLHPDYLSIP-ETIORIOLDIERRLDALELRGVELEKRLRAEGDAEDSLWY	579		
QY	247	EEPVLPQSSRLTLTKIRCADLY-RLPVRMSERPLQVNVDMANHLGVSPNRILL--FG	302		
Db	580	DMPWLHHEKOLL--LRQSELMYVSKARLEEQDQIDIGELRRRLMAKEALKLSQERRR	636		
QY	303	EESELPSTAPPTLKLGVADILDCVVLASSSEATETSOELRLVQGEKQHMLEISLSPDS	362		
Db	637	EEEL-----LEQVSTVND-----RSDIVSLIDERLADQ--EEDQMLRDMT-----	676		
QY	363	PLKVLMSHYEAMGLSGHKLSP	384		
Db	677	-----EKTLGLQKKSKF	688		
RESULT	11				
Q9CS98	ID	Q9CS98	PRELIMINARY;	PRT;	484 AA.
AC	Q9CS98	Q9CS98	PRELIMINARY;	PRT;	484 AA.
Db	01-JUN-2001	(TREMBLrel. 17, Created)			
Db	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
Db	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			

DE 3930402D05RIK PROTEIN (FRAGMENT).  
GN 3930402D05RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
RX MEDLINE=210836560; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gotojiori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glissl C., King B., Kochia H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
RA Blake J., Boffelli D., Botjuna N., Carrinci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita K., Gariboldi M.,  
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima Y., Mazzarelli J., Mombearts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilmink L.,  
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK014457; BAB29364.1; --  
DR MGD; MGI:1921269; 3930402D05RIK.  
SQ NON\_TER 1  
FT 1  
SEQUENCE 484 AA; 55413 MW; 328371FF52141F54 CRC64;

Query Match	Best Local	Similarity	6.1%	Score	128.5:	DB	11:	Length	484:
Matches	89:	Conservative	63:	Mismatches	158:	Indels	121:	Gaps	17:
QY	2	AEPLRGPRGRSGGAGARRAGA	-----	RGRCPAROSP	-----	35			
DB	103	AEROREDRRRSGCKEKERVKDRDRDRDKGRDRERRK	-----	SKNGENTDRPPREKSRDAKPEK	-----	162			
QY	36	-----	ARLLPVLVLDVSDSEEVLEVADPVEYVAKLPAPAKP	-----	-QGD	78			
DB	163	KSSSGEISRKISDGSFVKVAKEMEAD	-----	-TSVQASR-SSTLKPSKRRSKHSLEGD	-----	214			
QY	79	SDSDSEGAEGCAPRTLVRRRR	-----	-RLLDPGEAPVPVYSGKVOSLUI	-----	128			
DB	215	SPSDAEVEA-CPAGQDEPVEVMEVAPSELPSLRKIRIPRGASARAPPRVKKNOBESTERTLV	-----	273					
QY	129	PNNSLLKLCBPEDDEADLTNSSGSSPSEDDA	-----	-LPSSSPRRKTLRKCKCEKKEKKMEEPF	-----	186			
DB	274	VDRSSSGKTVSSV	-----	-IIDSONSDNEDDEQFVEVEAP	-----	313			
QY	187	DODISPLPPOPSRMRKHTTEALOKLRVKNRKLQDLRSCLSP	-----	-KOHOSPALOST	-----	242			
DB	314	DIDWVP	-----	-SGELEDEBKHHGVLKILFTKKNQYKYLQOSLKGKEKERSLIFESAKKREKD	-----	370			
QY	243	VVLVGPVLPPOSSRLFTLKIRCRADLVLRPRMSPELPNOVVDHANNHLCVSPNRLLLFG	-----	302					
DB	371	IYSKIEIKLRV	-----	-IQTLOCKSAI	-----	416			
QY	303	ESELSPTATPTSTLKGVDIIDCVVLASSSEATETSOELRLR	-----	-VOGKEKH	-----	351			
DB	417	ENRQIAEALSGE	-----	-QSITDSAVEPLKAELESEEQDIRQOQDKICAVKANILKNEKI	-----	471			
QY	352	OMLEISLSPDS	-----	362					
DB	472	OKMVSHINILSS	-----	482					

RESULT 12  
Q40363 PRELIMINARY: PRT: 635 AA.  
AC 040363:  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE NDM1 PROTEIN.  
RN NDM1.  
OS Medicago sativa (Alfalfa).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eucosids I; Fabales; Fabaceae; Papilionoideae; Medicago.  
OX NCBL\_TaxID=3879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96361876; PubMed=8721748;  
RA Boegre L., Jonak C., Mink M., Meskiane I., Traas J., Ha D.T.C.,  
RA Swoboda I., Plank C., Wagner E., Heberle-Bors E., Hirt H.;  
RT "Developmental and cell cycle regulation of alfalfa nucmst, a plant  
RT homolog of the yeast Nsr1 and mammalian nucleolin";  
RL Plant Cell 8:417-428(1996).  
DR EMBL: X88845; CAA61298.1;  
DR Mendel: 15257; Medsa: 2406; 15257.  
DR InterPro: IPR000504; RRM.  
DR Pfam: PF00076; rrm: 2.  
DR PROSITE: PSS0102; RRM: 2.  
DR SMART: SM00360; RRM: 2.  
SO SEQUENCE 635 AA; 67128 MW; DF17796FE52220F5 CRC64;

Query Match 6.1%; Score 128.5; DB 10; Length 635;  
Best Local Similarity 20.8%; Pred. No. 0.28;  
Matches 92; Conservative 75; Mismatches 183; Indels 93; Gaps 20;  
QY 24 ARGCRPARGSPA-----RLIPDVLVDLVD---SDEEVLEVADVEVPVAPRLPAPAK 74  
DB 80 AKAPAP-SKRTPAKGNVKAQOPETSESDSDISDEEVKPAKAVPSKNGSAPVK 138  
QY 75 -----PEODSDSDEGAEGPAGAPRTLVRRRRRLDDEGAPVAVVYSGVQSSINLIP 129  
DB 139 KDDESEEDSDSESSDEDDKPPAKAVPSKN-----GSAPAKKDDSEEDSDSDOE 191  
QY 130 DNSSLILKLPSE-----PEDEADLTNSGSPSEDDALPS-----GSPWRKKLKKKE 176  
DB 192 DEKPAKAVASSKNGSVSAKKDDSDSDSESDDEKPPAKAKSKVNSAPTAKAASSDE 251  
QY 177 KEKKMEEPDDDISPLPPSS--RNSSRKHTALQKLEVNKRLQDLRCLSPKQHQSPA 235  
DB 252 ESDSESDSDE--DEDAKPVSKPAAVAKSKKSDSDDEDDSDSDDEKPPVASKKEVSESE 309  
QY 236 LGSTDEEVVLVEBPVLPQSRLEFTLTKIRCADLVRLPVKMSPELQNVV-----IHANHL 290  
DB 310 SDSSDDDHKM---NIDKSDSDSESESEDEPLTKPQKIKDVE--MVAGKSGKKAAPT 365  
QY 291 GVSPPNILLFGESELSPTATSTLKL-----GVADIICVVIASSSE----- 333  
DB 366 PATPN-----ETSGSKLTFVGNLSFDVKRSDIENFPQGCYVVD--VRLASGDDVFG 417  
QY 334 -----ATETSOELRLVQGRK-KHQMELSL-----SPDSPLVLKMSHYEANGLSG 379  
DB 418 FGHEVATAEAAQSALEMNGOELHRLRLDLAREGAFTPNN-----NSNYSAGSGGGRG 472  
QY 380 HKLSFFPDG--TKLSGKELPADL 400  
DB 473 QSGTVEVVRGFDKNLGEDETRAKL 495

RESULT 13  
ID 061708 PRELIMINARY: PRT: 712 AA.  
AC 061708:  
DT 01-AUG-1998 (TReMBLrel. 07, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)  
DE R119.4 PROTEIN.  
GN R119.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBL\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Clarke K., Bauer C., O'Neal D.;  
RT "The sequence of C. elegans cosmid R119.";  
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF063007; AAD47125.1;  
SO SEQUENCE 712 AA; 77131 MW; 5858A2367A050ABA CRC64;

Query Match 6.1%; Score 128.5; DB 5; Length 712;  
Best Local Similarity 23.0%; Pred. No. 0.33;  
Matches 78; Conservative 33; Mismatches 111; Indels 117; Gaps 14;  
QY 7 GRGPRS-----RCGRGARRRGARGCPRARQSPARLIPDTVLVDLVS---D 52  
DB 140 GAAPRSTRGGEKPYVGRGGRG-----GARGYSRAVAPSALEPDATFADLDENQTKVD 194  
QY 53 EEVLEVADVE-----VVARLPAPAKPEQSDSDSEGAEGPAGAPRTVRRRRRL 106  
DB 195 TYTEVQPPVEESVTATVPTPSAPAPI-----SFAAANAARKEALKKHOQNP 244  
QY 107 DPEGAPVAVVYSGVQSSINLIPDNSSLILKLPSEPEDEADLTNSGSPSEDDALPSPGSP 166  
DB 245 QPAPAP-----RRSLSPQP-----LPSVAP 265  
QY 167 WRKK-----LKKCKEKKMEEPDDDISPLPPSSRNSRKHTALQKLEVNKRL 219  
DB 266 VKEEPAAPVFEPEPESHQPEKEGDFQONESSVLAEEQTPNVSTHHDENVOQS----- 317  
QY 220 QDLRCLSPKQHQSPAQ--STDEEVVLVEGPVLPQSRLEFTLTKIRCADLVRLPVKMSBP 278  
DB 318 -----TPPPNAWTQQLKTDIGLIGLSAPGIGLS-----PIPSAAP 353  
QY 279 LQNVYDHMANHLG-VSPNRL-LFGESELSPT-ATPST 314  
DB 354 VOIIPDPGEVGTAPTATNIDHSYSGFEVAPSPQPLST 392

RESULT 14  
ID 09VAJ5 PRELIMINARY: PRT: 755 AA.



Search completed: April 22, 2002, 10:11:59  
job time: 294 sec

---

This Page Blank (us30)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 22.6 Seconds  
(Without alignments)  
1350.363 Million cell updates/sec

Title: US-09-617-923-2  
Perfect score: 2099  
Sequence: 1 MAEPLRGGRSGRGAR.....GKELPADLGESGLIEYWG 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/genesqp/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/genesqp/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/genesqp/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/genesqp/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/genesqp/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/genesqp/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/genesqp/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/genesqp/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/genesqp/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/genesqp/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/genesqp/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/genesqp/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/genesqp/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/genesqp/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/genesqp/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/genesqp/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/genesqp/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/genesqp/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/genesqp/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/genesqp/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/genesqp/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/genesqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	18	AAW34091 Mouse NF-AT intera
2	2099	100.0	412	20	AAW08331 Mouse NIP45 protei
3	1399.5	66.7	419	20	AAW08330 Human NIP45 protei
4	623	29.7	138	22	AAW94289 Human protein sequ
5	132	6.3	586	18	AAW10423 Cytomegalovirus UL
6	126.5	6.0	676	22	AAW64373 Human hepatome cel
7	124.5	5.9	518	21	AAW53322 Human colon cancer
8	124.5	5.9	555	22	AAW53322 Human colon cancer
9	124.5	5.9	671	21	AAW99426 Human PRO1604 (UNQ
10	124.5	5.9	671	22	AAW66175 Protein of the lmv
11	123	5.9	950	20	AAW33298 Human membrane spa

12	122.5	5.8	1780	19	AAW53863 Human gravin polyp
13	122.5	5.8	1780	21	AAW15380 Human gravin prote
14	121	5.8	2971	21	AAW41231 Human ORF995
15	121	5.8	2972	22	AAW50363 Human SRCAP. Homo
16	121	5.8	3118	22	AAW50362 Human SRCAP. Homo
17	120.5	5.7	565	19	AAW61247 Streptococcus pneu
18	120.5	5.7	1881	21	AAW44506 Streptococcus pneu
19	120	5.7	669	19	AAW37483 Mouse liver cancer
20	118.5	5.6	508	21	AAW77945 A. thaliana enviro
21	118.5	5.6	846	21	AAW771057 Human membrane tra
22	118	5.6	576	20	AAW49541 Human PRO201 (Nsp1
23	118	5.6	576	20	AAW49545 Human Nsp1 mutant
24	118	5.6	576	20	AAW49546 Human Nsp1 mutant
25	118	5.6	576	20	AAW06477 Human tumour-assoc
26	118	5.6	576	21	AAW93684 Amino acid sequenc
27	118	5.6	576	21	AAW51932 Human PRO201 prote
28	118	5.6	576	21	AAW51939 Human PRO201 prote
29	118	5.6	987	22	AAW84633 Human protein sequ
30	117.5	5.6	815	22	AAW25561 Human CENP-C anti
31	117	5.6	979	14	AAW34783 Human tumour suppr
32	117	5.6	2663	22	AAW98612 Human Actinus L pro
33	115.5	5.5	1341	21	AAW85857 Human Nsp1 mutant
34	115	5.5	576	20	AAW49547 Peptide #3033 enco
35	114.5	5.5	92	22	AAW16599 Peptide #3120 enco
36	114.5	5.5	92	22	AAW29083 Peptide #2994 enco
37	114.5	5.5	92	22	AAW04312 Human protein sequ
38	114.5	5.5	454	22	AAW95394 Human striated mus
39	113.5	5.4	661	19	AAW77048 Human cell signal
40	113.5	5.4	661	21	AAW70078 Human normal bladd
41	113	5.4	519	21	AAW44247 Ubiquitin-like dom
42	113	5.4	605	22	AAW12251 Maize protein enco
43	113	5.4	783	20	AAW60344
44	112.5	5.4	95	20	AAW87985
45	112.5	5.4	610	12	AAW10923

#### ALIGNMENTS

RESULT	1
AAW34091	standard; Protein; 412 AA.
ID	AAW34091.
XX	AAW34091.
AC	AAW34091.
XX	18-MAY-1998 (first entry)
DF	XX
XX	XX
DE	Mouse NF-AT interacting protein 45.
XX	XX
KW	NF-AT Interacting Protein 45; NIP45; Yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	Interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation.
XX	XX
OS	Mus sp.
XX	XX
FT	Key
FT	Region
FT	Location/Qualifiers
FT	6..37 /note= "highly basic region"
PN	W09739721-A2.
XX	XX
PD	30-OCT-1997.
XX	XX
PF	23-APR-1997; 97WO-US06708.
XX	XX
PR	25-NOV-1996; 96US-0755592.
PR	23-APR-1996; 96US-0636602.
PR	25-NOV-1996; 96US-0755584.
XX	XX
PA	(HARD ) HARVARD COLLEGE.
XX	XX
PI	Glimcher LH, Ho I, Hodge MR;

```
XX WPI, 1997-535556/49.
DR N-PSDB; AAT93045.
XX
XX Production of cytokine(s) associated with Th2-type helper T cells -
PT particularly for controlling development of Th1 and Th2 cells for
PT treatment of allergy, autoimmune disease etc.
XX
XX Claim 31; Fig 11; 151pp; English.
XX
XX This is the amino acid sequence of the mouse NF-AT Interacting Protein 45
CC (NIP45). The gene sequence was isolated by using a yeast two-hybrid
CC detection system for proteins that interact with the NF-AT Rel homology
CC domain (RHD). The assay used, as a "bait", a 900 bp fragment of the
CC murine NF-ATP encoding the region spanning amino acids 228-250. NF-AT is
CC a multisubunit transcription complex containing a cyclosporin A sensitive
CC cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC of transcription factors. The screen was carried out on a cDNA library
CC prepared from the murine T cell line D10. One class of proteins, NIP45
CC designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC can be used in a claimed method to inhibit or stimulate production of
CC NF-AT family protein, particularly interleukin-4 (IL-4) and the treated
CC cells may be administered to control development of Th1 or Th2 cells by
CC modulating the production of a T-helper type 2 associated cytokine.
CC Especially the method is useful in the inhibition of Th2 in allergy,
CC cancer or infections, and promotion of Th2 in autoimmune disease and
CC transplantation.
XX
XX Sequence 412 AA:
SQ
Query Match 100.0%; Score 2099; DB 18; Length 412;
Best Local Similarity 100.0%; Pred. No. 9.5e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGPRSRGCGARGARGARCGRCPAROSPARIIPDTVLVLDVSDSEVLEVAD 60
DB 1 meepilrgprsrgrgarrargarrargrparqsparrilipdvlvldvsdsdeevleavd 60
QY 61 PVEVPVAPRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
DB 61 pvevpvparlpapakpeqdsdsdsegaegpagaprtlvrrrrrllldpgeapvvpvysgk 120
QY 121 VOSSLNLIPDNSSLKLCGSEPEDEADLTNCGSPSEDDALPSGSPWKKLKKCKEKEK 180
DB 121 vosslnltpdnssllklcgspepedeadltncgspseddalpsgspwkkllkkckeek 180
QY 181 KMEEPDODISPLPOPSRNKSRKHTKREVALOKLREVNKRLODDRSCSPKOHOSPALOSTD 240
DB 181 kmeepdodisplpopsrnrksrkhteaqlkirevnkrloddrscspkqhspalagstd 240
QY 241 DEVVIVEGVPVLPOSSRLFTLKITRCRADLVRLPYRMSEPLQNTVVDMANHLGVSPNRILL 300
DB 241 devvivegvpvlpssrflftklitrcradlvrlpyrmseplqnvvdhmanhlgvspnrilll 300
QY 301 FGSELSPTATPSTLKLGVADITDCVYLAASSSEATETSOELRLRVQGEKHKHMLEISLSP 360
DB 301 fgseelsptatpstlklgvaditdcvylasssseatseltrlyvgkexkhmleislsp 360
QY 361 DSPILKVMHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 412
DB 361 dspilkvlmshyeamglsghklsfffdgtrklsgkelpadlgiesgdliewwg 412
RESULT 2
AA08331
ID AAY08331 standard; protein; 412 AA.
XX
XX AAY08331;
AC
XX
XX 13-JUL-1999 (first entry)
DT
XX
XX Mouse NIP45 protein.
```

```
XX NIP45; mouse; transcription activator; IL-4; interleukin-4;
KW NF-AT interacting protein-45; modulator; treatment; inflammation;
KW autoimmune disease; HTP screening; drug testing; allergic therapy;
KW T-cell dependent response; cytokine; diagnostic; immunosuppressant.
XX
XX Mus musculus.
XX
XX WO9921993-A1.
XX
XX 06-MAY-1999.
XX
XX 21-OCT-1998; 98WO-GB03141.
XX
XX 24-OCT-1997; 97GB-0023388.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Liu D, Zhao J, Zhou H;
XX
XX WPI, 1999-312964/26.
XX
XX Polynucleotides encoding human NF-AT interacting protein (NIP45)
XX
XX Disclosure; Fig 4; 84pp; English.
XX
XX This invention describes a novel human NF-AT interacting protein-45
XX NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
XX interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
XX to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
XX autoimmune disease. NIP45 is a good candidate target for HTP screening
XX and/or testing system for drugs that will alleviate T-cell dependent
XX autoimmune and allergic responses, and for cytokine-based therapies of
XX chronic disease. Expression of NIP45 can be inhibited, and IL-4
XX expression modulated, by administering an antisense molecule. Antibodies
XX against NIP45 and primers can be used in diagnostic assays. Discovery of
XX a human NIP45 is advantageous in that it provides the ability to control
XX IL-4 transcription, which is of importance for anti-inflammation and
XX immunosuppressant drug development.
XX
XX Sequence 412 AA:
SQ
Query Match 100.0%; Score 2099; DB 20; Length 412;
Best Local Similarity 100.0%; Pred. No. 9.5e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGPRSRGCGARGARGARCGRCPAROSPARIIPDTVLVLDVSDSEVLEVAD 60
DB 1 meepilrgprsrgrgarrargarrargrparqsparrilipdvlvldvsdsdeevleavd 60
QY 61 PVEVPVAPRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPVYSGK 120
DB 61 pvevpvparlpapakpeqdsdsdsegaegpagaprtlvrrrrrllldpgeapvvpvysgk 120
QY 121 VOSSLNLIPDNSSLKLCGSEPEDEADLTNCGSPSEDDALPSGSPWKKLKKCKEKEK 180
DB 121 vosslnltpdnssllklcgspepedeadltncgspseddalpsgspwkkllkkckeek 180
QY 181 KMEEPDODISPLPOPSRNKSRKHTKREVALOKLREVNKRLODDRSCSPKOHOSPALOSTD 240
DB 181 kmeepdodisplpopsrnrksrkhteaqlkirevnkrloddrscspkqhspalagstd 240
QY 241 DEVVIVEGVPVLPOSSRLFTLKITRCRADLVRLPYRMSEPLQNTVVDMANHLGVSPNRILL 300
DB 241 devvivegvpvlpssrflftklitrcradlvrlpyrmseplqnvvdhmanhlgvspnrilll 300
QY 301 FGSELSPTATPSTLKLGVADITDCVYLAASSSEATETSOELRLRVQGEKHKHMLEISLSP 360
DB 301 fgseelsptatpstlklgvaditdcvylasssseatseltrlyvgkexkhmleislsp 360
QY 361 DSPILKVMHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGIESGDLIEWWG 412
DB 361 dspilkvlmshyeamglsghklsfffdgtrklsgkelpadlgiesgdliewwg 412
```



CC CMV infection.

Human hepatome derivative growth factor polypeptide sequence and its preparation process

XX

PS Claim 2; Pages 15-16 (Disclosure); 27pp; Chinese.

XX The present sequence is the protein sequence of III-type human hepatome  
CC cell derived growth factor (HGF3). HGF3 is the homolog of mouse HRP-2.  
CC Also disclosed are the application of the HGF3 coding sequence and  
CC protein and the process for preparing the coding sequence and protein.  
XX

SQ Sequence 676 AA:

Query Match 6.0%; Score 126.5; DB 22; Length 676;  
Best Local Similarity 21.9%; Pred. No. 0.025;  
Matches 103; Conservative 50; Mismatches 149; Indels 169; Gaps 23;

```
QY 8 RGRSRGCGARRARCARCPRAROSPARLIPDYLVLY-----SDSD 52
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 213 repr-rgprlggrtkk-----kapsasdsdskadsgakpepvamaraasssssssd 267
QY 53 EEWLE-----VADPEVVARLPAPAKP-----EQSDSD-----SEGAAGPAGAPRT 96
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 vsvkkpprgtkpaekpdpkprgrtkprrpssssdssddevdrisewkrrdearrrel 327
QY 97 LVRRRR-----RL-----LDPEAPVNVYSGKVOSSINLIPDSSSLK 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 earrtrgeeeelrrlregekeekerrerradrgae-----rgsgsgsgdelreddepvk 382
QY 137 -----LCPSEED--EADLTNSGSSPSDDALPSGSPRRKRLRCKEKEEK 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 383 krgtkgrgrgppsssdsepeaeereakksakkkpsssteapkpqke--kvrpreeq 440
QY 182 MEEFPDQDISPLPQSSRNKSRKHTALQKLRVNRRLDLSRCLSPKQHSPPALOSTDD 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 441 q-----akpvttertkrsegfmdtkvkekke-----psveek--lqklhs 480
QY 242 EYVLVGPVLPQSSRLFTLKI-----RCRADLVRLPVRM--SEPLQNVVDHMANHLGVS 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 481 el-----kftakvspdkrcinalaeelgtlqvtsqlqkntdvva----- 521
QY 294 PNRILLFGESELSPTATPSTLK-----LGVADIIDCVVIASSSEATFNSOELRLRVCK 348
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 522 -----tlkkirrykankdyme-----kaeaytrtksrviqp 553
QY 349 EKHQMLEISLSPDPLKVLMSHSEAMGLSGHKLSEFFDGTKLSGKELPAD 399
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 554 ktaevqkv-----kagmekakeeklagaela-----geelagaeapqe 593
```

RESULT 7

AAB53322 standard; Protein; 518 AA.

XX AAB53322;  
XX  
XX  
XX 09-MAR-2001 (first entry)  
XX  
XX Human colon cancer antigen protein sequence SEQ ID NO:862.  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX identification; cytostatic; cardioactive; neuroprotective; vulnery;  
XX immunomodulatory; muscular; gynaecological; gastrointestinal;  
XX nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;  
XX neutral disorder; immune system disorder; muscular disorder;  
XX reproductive disorder; gastrointestinal disorder; renal disorder;  
XX infectious disease; cardiovascular disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200055351-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587534/55.

DR N-PSDB; AAC98079.

XX Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -

PS Claim 11; Page 1414-1416; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53324 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnery, nephrotoxic, antineoplastic, antibacterial, and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins  
CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders, immune  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.  
XX

SQ Sequence 518 AA:

Query Match 5.9%; Score 124.5; DB 21; Length 518;  
Best Local Similarity 23.1%; Pred. No. 0.026;  
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;

```
QY 8 RGRSRGCGARRARCARCPRAROSPARLIPDYLVLY-----SDSD 52
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 repr-rgprlggrtkk-----kapsasdsdskadsgakpepvamaraasssssssd 151
QY 53 EEWLE-----VADPEVVARLPAPAKP-----EQSDSD-----SEGAAGPAGAPRT 96
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 152 vsvkkpprgtkpaekpdpkprgrtkprrpssssdssddevdrisekrrdearrrel 211
QY 97 LVRRRR-----RL-----LDPEAPVNVYSGKVOSSINLIPDSSSLK 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 212 earrtrgeeeelrrlregekeekerrerradrgae-----rgsgsgsgdelreddepvk 266
QY 137 -----LCPSEED--EADLTNSGSSPSDDALPSGSPRRKRLR----- 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 267 krgtkgrgrgppsssdsepeaeereakksakkkpsssteapkpqkekvrrpreeq 326
QY 173 --KCKEKEKKKEEPP--DDDISPLPQSSRNKSRKHTALQKLR--EVNRRLQ-----DLRS 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 327 kvvttertkregfmdtkvkekkepsveek-----lqllhaeikalkavdspdkvr 379
QY 225 CUS-----PKQHSPPALOSTDDEVLYE-----GVLDPQSSRLFTLKRADLV 270
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 380 cinalaeelgtlqvtsqlqkntdvvaltkirrykankdymeaeayv-----r 429
QY 271 LPRMSEPLQNVVDHMANHLGVSPPRILLFGESELSPTATP 312
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 430 lksrvlqpktaevqkv-nkagmekakeeklagaelagaeap 470
```

RESULT 8

AAG75073 standard; Protein; 555 AA.

XX AAG75073;  
XX  
XX  
XX

Query Match	5.9%; Score 124.5; DB 22; Length 555;
Best Local Similarity	23.1%; Pred. No. 0.029;
Matches	93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;
01	Sequence 555 AA;
02	
03	
04	
05	
06	
07	
08	
09	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

Accession	Protein Name	Sequence
Oy	173 --KKEKEKEKKKEEP-DQDSDPLPQSSRNKSKNHTALQKL-R-EVVKRLQ-----DLRS 222	98US-0098716. 98US-0098746. 98US-0098750. 98US-0098801. 98US-0098821. 98US-0098843. 98US-0098936. 98US-0099596. 98US-0099599. 98US-0099602. 98US-0099642. 98US-0099741. 98US-0099754. 98US-0099763. 98US-0099792. 98US-0099808. 98US-0099812. 98US-0099815. 98US-0099816. 98US-0099816. 98US-0100385. 98US-0100388. 98US-0100390. 98US-0100584. 98US-0100627. 98US-0100661. 98US-0100664. 98US-0100683. 98US-0100684. 98US-0100710. 98US-0100711. 98US-0100919. 98US-0100930. 98US-0100848. 98US-0100849. 98US-0101014. 98US-0101066. 98US-0101071.
Db	327 kvpkvtrkrsegsfmsdrkvekkkxpsveek-----lqklnselkfkvdspdvkr 375	
Oy	225 CUS-----PKOHSPALQSDTDEVLVE-----GPIVPGSSRLFTLKIRCRADLV 270	
Db	380 cnaaleelgtlgtvtsqllqkntdvaatlklirrykankdvmeakaeyt-----r 422	
Oy	271 LPVRMSEPLQNVVDHMANHLGVSPNRILLFGESLSTATP 312	
Db	430 lksrlylqpkleavqkv-nkagmekekaeklagelagaeap 470	
RESULT 9		
ID	AAI99426 standard; Protein; 671 AA.	
XX	AAI99426;	
XX	08-AUG-2000 (first entry)	
DE	Human PRO1604 (UNQ785) amino acid sequence SEQ ID NO:308.	
XX	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis; transmembrane; secretion; immunoadhesion; pharmaceutical; screening.	
XX	Homo sapiens.	
XX	WO200012708-A2.	
XX	09-MAR-2000.	
XX	01-SEP-1999; 99WO-US20111.	
XX	01-SEP-1998; 98US-0098716.	
PR	01-SEP-1998; 98US-0098746.	
PR	01-SEP-1998; 98US-0098750.	
PR	02-SEP-1998; 98US-0098801.	
PR	02-SEP-1998; 98US-0098821.	
PR	02-SEP-1998; 98US-0098843.	
PR	02-SEP-1998; 98US-0099536.	
PR	09-SEP-1998; 98US-0099596.	
PR	09-SEP-1998; 98US-0099599.	
PR	09-SEP-1998; 98US-0099602.	
PR	09-SEP-1998; 98US-0099642.	
PR	10-SEP-1998; 98US-0099741.	
PR	10-SEP-1998; 98US-0099754.	
PR	10-SEP-1998; 98US-0099763.	
PR	10-SEP-1998; 98US-0099792.	
PR	10-SEP-1998; 98US-0099808.	
PR	10-SEP-1998; 98US-0099812.	
PR	10-SEP-1998; 98US-0099815.	
PR	10-SEP-1998; 98US-0099816.	
PR	15-SEP-1998; 98US-0100385.	
PR	15-SEP-1998; 98US-0100388.	
PR	15-SEP-1998; 98US-0100390.	
PR	16-SEP-1998; 98US-0100584.	
PR	16-SEP-1998; 98US-0100627.	
PR	16-SEP-1998; 98US-0100661.	
PR	16-SEP-1998; 98US-0100664.	
PR	17-SEP-1998; 98US-0100683.	
PR	17-SEP-1998; 98US-0100684.	
PR	17-SEP-1998; 98US-0100710.	
PR	17-SEP-1998; 98US-0100711.	
PR	17-SEP-1998; 98US-0100919.	
PR	17-SEP-1998; 98US-0100930.	
PR	18-SEP-1998; 98US-0100848.	
PR	18-SEP-1998; 98US-0100849.	
PR	18-SEP-1998; 98US-0101014.	
PR	18-SEP-1998; 98US-0101066.	
PR	18-SEP-1998; 98US-0101071.	











Query Match	Score	DB	Length
5.8%;	121;	21;	2971;

Best Local Similarity 21.28; PRed.NO. 0.01;  
Matches 102; Conservative 41; Mismatches 173; Indels 166; Gaps 18;

OY		6	RGRGRRSGCGAGARRAGATG-----		-GC-----	-PRA	31				
Dd		2125	kakapeipgtctvserlligaaetcganhhprvsahqtstlttpprcsparevrprap				2184				
OY		32	ROSFA--RLIPDTVLVDVSDSEVELEVADPEVY-PVARLPAPAKPEDSDSDSEGAA	87							
Dd		2185	rftpasapaipalvpvy---sapypisapnptitltpvhillpspppsqilppcsspact	2241							
OY		88	EGRPAGAP-----RTLYRRRRRRLDPGEAPVPYPYSGKVGSSLNLIPD-----	130							
Dd		2242	pppactpppahhtpppaqctivrtcsspllllp---psvplsasavtnlplqlfpeaelcaqa	2298							
OY		131	-----NSLKLKLCPSF-----PEDEADLTNGSSSPS-EDDALP	162							
Dd		2299	laspeslelaavaasetstslvvpkdllpvaveellpvseknlsiltspapsiltleagslp	2358							
OY		163	SGSWRRKKLRKCKEKEKNHE-----FPDDDISPLPPQSSRN----KSRRHTFALOKL	212							
Dd		2359	nqq-----eqeapdsaeqtlvtlvipegeelplcvsesngtelppsaaedepiqe	2408							
OY		213	REVNKRLLQDLRSCLSPQHOSPALQSTDEEVYL-----VGGPYLPQSSRLFUTKI	262							
Dd		2409	leadrtseelreaktcrpspekpgelvtaveaapstssatsatsspqpaparptr-----	2463							
OY		263	RCRADLVRLPRMSEPIONVVDMANHLGVSPRIILLFGESEL-----SPNAPDSITUKL	317							
Dd		2464	rttsadveitrggvtgrppgpqpkvlrkl---pgrlvtvveekelvrrrrrqrgaastlvp	2520							
OY		318	GVAAILDCVWLASSSEATEFTSQELRLRVOGKEKHQMLETISLSPDSDPLKYILASHYEEMGL	377							
Dd		2521	gvs-----etsaspagspsvrysmagpesappi	2546							
OY		378	SG	379							
Dd		2547	gg	2548							
RESULT_15											
AAB50363	ID	AAB50363	standard; protein; 2972 AA.								
XX	XX	AAB50363;									
XX	XX	12-MAR-2001	(first entry)								
XX	XX	Human SRCAP.									
XX	XX	Human: SRCAP, Snf2 related CBP activator protein; antiviral; CREB;									
KW	KW	CAMP regulatory element; CREB binding protein; CBP; ATPase;									
KW	KW	transcription activation; DEAD box RNA dependent helicase;									
RM	RM	adenoviral DBP protein; beta-actin; nuclear receptor; viral infection.									
XX	XX	Homo sapiens.									
OS	OS	M0200073467-A1.									
XX	XX	07-DEC-2000.									
XX	XX	25-MAY-2000; 2000MO-US14719.									
XX	XX	27-MAY-1999; 99US-O136620.									
XX	XX	25-MAY-2000; 2000US-0579181.									
XX	XX	(TUSL-) UNIV SAINT LOUIS.									
XX	XX	Chrivia J, Yaciuk P;									
XX	XX										

DR WPI: 2001-061545/07.  
DR N-PSDB: AAC89860.  
XX  
XX Snt2 related cAMP regulatory element (CREB) binding protein (CBP)  
PI activator protein, capable of co-activating CREB binding protein,  
PI useful for modulating transcription and for affecting viral infection  
XX  
XX  
PS Claim 1; Page 86-94; 103pp; English.

The present sequence is an Sfr-related CREB (cAMP regulatory element) binding protein (CBP) activator protein (SRAP) polypeptide. It has ATPase activity and is capable of activating transcription. SRAP polypeptides are useful for activating transcription in a cell, for enhancing CREB (cAMP regulatory element) binding protein (CBP)-mediated activation of transcription in a cell, for treating a patient having a disease involving a function such as insufficient transcription of a gene, e.g. a gene mediated by CBP co-activation, DEAD box RNA dependent helicase, adenoviral DBP protein, beta-actin or a nuclear receptor affected by SRAP protein. Compounds that modulate SRAP function, such as antibodies, antisense molecules, polynucleotides or ribozymes, are useful for treating diseases mediated by SRAP-activated transcription, for example, infection by adenovirus, hepatitis C virus, human immunodeficiency virus type-1, Epstein Barr virus, cytomegalovirus or hepatitis B virus.

Query Match	5.8%	Score 121	DB 22	Length 2972
Best Local Similarity	21.2%	Pred. No. 0.67		
Matches 102; Conservative	41	Mismatches	173	Gaps 18

```

OY      6  RGCRGRSGCGGARRRAGAG-----RC-----PRA 31
Db      2125 kakapetpgtvtvserlllgataacgaanhvrsahbhtsttcttppctsparevrvpprap 2184a
OY      32  ROSPA--RLIPDTVLVDLVSDSEVLEVADPEVY-PVARKLPAPARKPEQSDSDSEGA 87
Db      2185  rptbasapaipalvpyvv--sawpvisapnbtillpvhlppppsq!ppcsspact 2241a
OY      88  EGPAGAP-----RTLVRRRRRRLDPGBAPVVPYSGVQSLMLTD----- 130
Db      2242  pppactpppactpppactctltvpsplllyp---psvplsaaytnlpjglrpeaelcga 2299a
OY      131  -----NSSLLKLCPSF-----PEDEADLTNMSGSSPS---EDDALP 162
Db      2299  laspeslelasvaasetssislvpkdl!pvaveilpvseknsltspasltleagshp 2356a
OY      163  SCSFWRRKKLRKCKCKEKKKME-----PPDDISLFLPPSSRN---KSKHNTALOKL 212
Db      2359  ngq-----eqeapdaegctltvlrpegeelr!cvssengle!lpsaasdeplqep 2408a
OY      213  REVNKRRLQDLRSCLSPKQHOSPALQSTNDPEVVL-----VSGPVLPQSSRLFTLKI 262
Db      2409  leadtseelteaktprpspekpqgelvtvaeapstssatspspgpaprpr----- 2463a
OY      263  RCRADLVRLPKRSEPLQNVNDMANHGLVSPRRILLFGESEL-----SPAPRSTLKL 317
Db      2464  rtsadevrlrqgtlrpqgppqkvllkl---pgrlvtvveekel!vqrrtqgaaastlv 2520a
OY      318  GVAADIDCVVLASSSEATENTSOELRLRVQGEKKNQMLETLSLPPSDPLKVIWMSHYEANGL 377
Db      2521  gvs-----elcsapgspsvmsgmgespspl 2546a
OY      378  SG 379
Db      2547  gg 2548

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 19.24 Seconds  
(without alignments)  
481.879 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099

Sequence: 1 MAEPLRGGRGPRGRCGRCARR.....GKELPADLGLESGDLIEWWG 412

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6CTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	2	US-08-755-584-2
2	2099	100.0	412	3	US-09-192-611-2
3	122.5	5.8	1780	1	US-08-769-309A-5
4	122.5	5.8	1780	3	US-08-994-570-5
5	120.5	5.7	565	4	US-08-961-083-218
6	113.5	5.4	661	2	US-08-795-868-14
7	112.5	5.4	95	4	US-09-100-802-10
8	111	5.3	897	1	US-08-095-737-4
9	111	5.3	897	2	US-08-480-145-4
10	111	5.3	897	2	US-08-477-389-4
11	109.5	5.2	1142	4	US-09-061-709-2
12	108.5	5.2	504	4	US-08-955-918C-7
13	108.5	5.2	505	1	US-08-631-200-2
14	108.5	5.2	505	1	US-08-630-592-4
15	108.5	5.2	505	1	US-08-714-991-4
16	108.5	5.2	505	2	US-08-829-553-2
17	108.5	5.2	505	2	US-08-922-267A-2
18	108.5	5.2	505	2	US-08-936-707A-2
19	108.5	5.2	505	2	US-08-936-706A-2
20	108.5	5.2	505	3	US-09-248-203-2
21	108.5	5.2	505	3	US-09-032-365A-4
22	108.5	5.2	505	4	US-08-812-824-3
23	108.5	5.2	505	4	US-09-406-071-2
24	108	5.1	459	1	US-08-630-592-2
25	108	5.1	459	1	US-08-714-991-2
26	108	5.1	459	3	US-09-032-365A-2
27	108	5.1	673	2	US-08-435-073A-6

28	108	5.1	1001	4	US-09-060-410-2	Sequence 2, Appl1
29	107	5.1	521	2	US-08-721-684C-2	Sequence 2, Appl1
30	107	5.1	521	2	US-09-005-970-2	Sequence 2, Appl1
31	107	5.1	521	4	US-09-407-715-2	Sequence 2, Appl1
32	107	5.1	570	3	US-08-826-246-2	Sequence 2, Appl1
33	107	5.1	570	3	US-08-944-495-2	Sequence 2, Appl1
34	107	5.1	570	3	US-09-126-640-7	Sequence 2, Appl1
35	107	5.1	570	4	US-08-925-588-2	Sequence 2, Appl1
36	106.5	5.1	723	2	US-08-548-159-5	Sequence 2, Appl1
37	106.5	5.1	1142	2	US-08-993-118-7	Sequence 7, Appl1
38	106.5	5.1	1142	3	US-08-845-528C-7	Sequence 7, Appl1
39	104.5	5.0	861	1	US-08-484-105-18	Sequence 18, Appl1
40	104.5	5.0	861	1	US-08-484-106-18	Sequence 18, Appl1
41	104	5.0	126	2	US-08-853-974-1	Sequence 1, Appl1
42	104	5.0	126	4	US-09-172-988-1	Sequence 1, Appl1
43	104	5.0	3248	1	US-08-353-700-1	Sequence 1, Appl1
44	104	5.0	3248	5	PCT-US95-16216-1	Sequence 1, Appl1
45	103.5	4.9	586	2	US-08-630-822A-70	Sequence 70, Appl1

## ALIGNMENTS

RESULT 1  
US-08-755-584-2  
; Sequence 2, Application US/08755584  
; Patent No. 5858711  
GENERAL INFORMATION:  
; APPLICANT: Gilmcher, Laurie H.  
; APPLICANT: Hodge, Martin R.  
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP4S AND METHODS  
; NUMBER OF INVENTION: OF USE THEREFOR  
; NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/755,584  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Giulio A. DeConti, Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: HUI-026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-755-584-2

Query Match 100.0%; Score 2099; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 1.8e-174;  
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEPLRGGRGPRGRCGRCARRCPARRSPARLIPDTVLVDVSDSEVLEVAD 60

```
|||||
Db 1 MAEPLRGPRGRGRCGARRGARGRCPRARQSPARLLIPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPARLPAPAKPEQDSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAVPVVYSGK 120
Db 61 PVEVPARLPAPAKPEQDSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAVPVVYSGK 120
QY 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
Db 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
QY 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSKQKOSPALQSTD 240
Db 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSKQKOSPALQSTD 240
QY 241 DEVVLVEGVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
Db 241 DEVVLVEGVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGESELSPTATPSTLTKLGVAADIIDCVYLAASSSEATETSOELRLRVQKREKHOMLEISLSP 360
Db 301 FGESELSPTATPSTLTKLGVAADIIDCVYLAASSSEATETSOELRLRVQKREKHOMLEISLSP 360
QY 361 DSPPLKVMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLSGDLIEWMG 412
Db 361 DSPPLKVMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLSGDLIEWMG 412

RESULT 2
US-09-192-611-2
; Sequence 2, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; APPLICANT: Hodges, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,611
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. Decontil, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-192-611-2
```

```
Query Match 100.0%; Score 2099; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1,8e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGPRGRGRCGARRGARGRCPRARQSPARLLIPDTVLVDVSDSEVLEVAD 60
Db 1 MAEPLRGPRGRGRCGARRGARGRCPRARQSPARLLIPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPARLPAPAKPEQDSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAVPVVYSGK 120
Db 61 PVEVPARLPAPAKPEQDSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAVPVVYSGK 120
QY 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
Db 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
QY 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSKQKOSPALQSTD 240
Db 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSKQKOSPALQSTD 240
QY 241 DEVVLVEGVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
Db 241 DEVVLVEGVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGESELSPTATPSTLTKLGVAADIIDCVYLAASSSEATETSOELRLRVQKREKHOMLEISLSP 360
Db 301 FGESELSPTATPSTLTKLGVAADIIDCVYLAASSSEATETSOELRLRVQKREKHOMLEISLSP 360
QY 361 DSPPLKVMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLSGDLIEWMG 412
Db 361 DSPPLKVMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLSGDLIEWMG 412

RESULT 3
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Naert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

TELEPHONE: 312-474-0300  
TELEFAX: 312-474-0448

?  
 ? PRIOR APPLICATION DATA:  
 ?  
 ? APPLICATION NUMBER:  
 ?

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 218:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-218

Query Match 5.7%; Score 120.5; DB 4; Length 565;  
Best Local Similarity 20.9%; Pred. No. 0.02;  
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRARSP-----ARLPDVLVDVSDSEVLEVADPYEV--PYARLP--APAKPE-Q 77  
DB 34 PTEESPKEEPKSEVAKPTDDTLPKVEBEKEDSAEPAPVEEVGEVSEKPEEKVAAKPEQ 93  
QY 78 DSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPYSGVQSS--LNLIPDNSSL 134  
DB 94 PEDKCAEESKVGQAGEP-VAPREDEKAPVEPKQEPAPPEEKAAVEETPKQESTPTDKAE 152  
QY 135 LALCP-----SEPEDADLTNGS--SPSEDDALPSCGFWRK 170  
DB 153 ETVPEPEETVNOSEIDPKVETPAVEKQETPEPEKVEQAGEVPAPREDQAPT-AVEPE 211  
QY 171 LKKCKEKEKEEPPDDDISPLPQSSRNKSKHTHEALQKLEVVKRLQDLSRCSKPKQ 230  
DB 212 KQPEVPEEKAVBE-----TPKREDKIKGIGTEPEYDK-SELNNQI-DKASSSPFD 261  
QY 231 HOSPALQSTDEVEVLEGPVLPQSSRLFTLTKRCRADLVRLPVRMSE--PLQNVVD---- 284  
DB 262 YSTASYNAL-----GVLETAQGYA-----SEVPKQEPVNSETKLTALDALNV 307  
QY 285 ---HANHLGVSPNRLLLFGESELSPTATPSTLKIGVADIIDCVLASSSEATETSQSL 341  
DB 308 DKTELNNLTADKTKYKEHYSDRSWONIQTEVYTKAEKVAANTD---AKQSEVNEAVEKL 363  
QY 342 RLVOGKEKHOMLEISLSPDSPKLVLMS 369  
DB 364 TATIE-----KVELS---EKPLITLTS 383

RESULT 6  
US-08-795-868-14  
Sequence 14, Application US/08795868  
Patent No. 5846773  
GENERAL INFORMATION:  
APPLICANT: Lee, Mu-En  
APPLICANT: Hsieh, Chung-Ming  
TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC  
TITLE OF INVENTION: AND STRAINED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,868  
FILING DATE: 06-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/494,577  
FILING DATE: 22-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/032001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 661 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-795-868-14

Query Match 5.4%; Score 113.5; DB 2; Length 661;  
Best Local Similarity 20.0%; Pred. No. 0.1;  
Matches 103; Conservative 55; Mismatches 131; Indels 227; Gaps 25;

QY 4 PLRGR---GPRRG---KGRARRAGAR----- 25  
DB 26 PARGHVHAAPPSHGACAARGHNRREAREYQVVALPGHRESRPOTPLSEASGRIMA 85  
QY 26 -GRCRRARQSPARLLIPDVLVDVSDSE--EVLVADPVEVPV-----ARLPAPAK 74  
DB 86 LGRSRLVRAAGSRIL-----DKLOFPEERRSRLERSDSPAPRLPWPVPLKRAARLEQPK 139  
QY 75 PEQSDSDSEGAEGPAGAPRTLVRRR-----RRLDPGEAPVVPYS-----GK 120  
DB 140 SERGAPWGTGPGASOELRAFPGSVAAERRRLFOOKAASLDERTRORSASDLELRFQOELGR 199  
QY 121 VOSSLN---LIPDNSSL--LKLCPSEPE-DEADLTNGSSPSSEDDAL----- 161  
DB 200 IRRSTSRRELVSHESLATLQRAPSPREPGPLFSRSTPKTSRAASPAAPPSPS 259  
QY 162 ---PSGSPWRKRLRKKC-----EKEKKMEFP-----DODISPLPQPS----- 197  
DB 260 AKPQDEPGRRSRGPAGTPEGBGPQDEVRRRDQFPLRSRAIOECSPVPPAPADPE 319  
QY 198 SRNKS---RKHTALQKL-----EVNKRLO----- 220  
DB 320 ARTKAPGRRKRRPPAQAVALFLPWATPGLGAAVPTQLEKNRAGPAERKRLRGPREDEGPW 379  
QY 221 ---DLRSLCSLQKHQ-----SPALQSTDEVYL-----VEGVLPQSSSLFTLKIRCRAD 267  
DB 380 GPWDRRGARQSGKGRARPTSELSSDSYSTVSAQEEPIEAPV----- 422  
QY 268 LVRLPVRMSEPLQNVVDHANHLGVSPNRLLLFGESELSPTATPSTLKIGVADIIDCV 327  
DB 423 -----FELPQNVV-----VAP-----GADVLLKCI 444  
QY 328 LAS-----SSEATETSQELRLRVQCKEKQML 354  
DB 445 TANPPQVSWHKDGSALRSEGRLLRLRAEG-ERHTLL 479

RESULT 7  
US-09-100-802-10  
Sequence 10, Application US/09100802A  
Patent No. 6294363  
GENERAL INFORMATION:  
APPLICANT: Madura, Kiran  
TITLE OF INVENTION: Methods and Compositions for the Rapid



;; TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components  
;; FILE REFERENCE: UMDNJ97-11  
;; CURRENT APPLICATION NUMBER: US/09/100,802A  
;; EARLIER FILING DATE: 1998-06-19  
;; NUMBER OF SEQ ID NOS: 17  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 10  
;; LENGTH: 95  
;; TYPE: PKT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Synthetic Sequence  
US-09-100-802-10

Query Match 5.4%; Score 112.5; DB 4; Length 95;  
Best Local Similarity 32.5%; Pred. No. 0.0085;  
Matches 25; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

OY 335 TETSQLRVRGKEKHOMLEISLSPDKVLMHYEAMGLSGHLSFFDGTLSGK 394  
DB 12 TENNDININKVAGOD-GSYVOFKIKRHTPLSKLMKAYCERQISMRQIRFRFDGPINET 70  
OY 395 ELPADGLSGDLIEYW 411  
DB 71 DTPPAOLEMEDEDTIVF 87

RESULT 8  
US-08-095-737-4  
;; Sequence 4, Application US/08095737  
;; Patent No. 5487979  
;; GENERAL INFORMATION:  
;; APPLICANT: DiFiore, Pier P  
;; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Knobbe, Martens, Olson & Bear  
;; STREET: 620 Newport Center Drive, Sixteenth Floor  
;; CITY: Newport Beach  
;; STATE: California  
;; COUNTRY: United States of America  
;; ZIP: 92660  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: IBM PC compatible  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/095,737  
;; FILING DATE: 19930722  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Israelsen, Ned A  
;; REGISTRATION NUMBER: 29,655  
;; REFERENCE/DOCKET NUMBER: NIH060.001A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 235-8550  
;; TELEFAX: (619) 235-0176  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 897 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-095-737-4

Query Match 5.3%; Score 111; DB 1; Length 897;  
Best Local Similarity 21.2%; Pred. No. 0.26;  
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

OY 61 PVEVPVAPRLP-----APA-KPED-----SDSDEGAAGC-----PAGAPRTL 97  
DB 202 PMSLPALVPSPSKRTVWVSPAERKAYDEIFLTKDKMDGVSGLEVRETFKLTGPSAL 261  
OY 98 VRRRRRLDPGAPVVPVVS-----GRVQ-----SINLPDMSLLK-----LC 138  
DB 262 LAH-----INSLCDTRGCGKLSKDPALAPHLI--NOKLTKGIDPPHSLT 304  
OY 139 PS--EPEDADLTN--SGSSPSED-----DALPSGSPFRKRLRKCKEKKMEFPD 187  
DB 305 PEMIPSDRSSLOKNITGSSPVADFSKIKELDTLNEIYVLQEKKNVDEDLKEKEDTVK 364  
OY 188 QDISPLPQSSRNKSRKHTALQKLRVNRLODLRSCLSPKQHQ-----SPALQSTDE 242  
DB 365 QRTSEV--QDLQDEYQRESINLQKLOKQOQVQELLGELDEQKAQLEQLOQVRRKCAEE 422  
OY 243 VVLV---EGVVLQSSKRLFTLK---ITCRADLYRLPYRMSEPLQNV-----VDHANHL 290  
DB 423 AOLISLKAETISOESQOISSYEEBELKARELSLQOETQOLESEVSGKAOLEPLOQHL 482  
OY 291 GVSPPNRIILFGSELSPTAT-----PSTLKLGVADITDCVVLASSEATEETSQ 339  
DB 483 QESQOETSSQMOMLEKDELTDNNQSMSSPOSVLYNGATDY--CSLTSSETANFNE 540  
OY 340 --ELRLVQGEKHKHOMLEISLSPD-SPLKVLMSHYEAMGLSGH-KLSFFDGTLSGKE 395  
DB 541 HAQGNMLESEPTQOESSVNSPEIAPSDV--TDESAVTVAGNEKVTPRFDDKHSKEE 598  
OY 396 LPADGLSGDLIE 409  
DB 599 DP--FNVESSLTD 610

RESULT 9  
US-08-480-145-4  
;; Sequence 4, Application US/08480145  
;; Patent No. 5717067  
;; GENERAL INFORMATION:  
;; APPLICANT: DiFiore, Pier P  
;; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Knobbe, Martens, Olson & Bear  
;; STREET: 620 Newport Center Drive, Sixteenth Floor  
;; CITY: Newport Beach  
;; STATE: California  
;; COUNTRY: United States of America  
;; ZIP: 92660  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: IBM PC compatible  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/480,145  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/095,737  
;; FILING DATE: 22-JUL-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Israelsen, Ned A  
;; REGISTRATION NUMBER: 29,655  
;; REFERENCE/DOCKET NUMBER: NIH060.001A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 235-8550  
;; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 897 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-480-145-4

Query Match 5.3%; Score 111; DB 1; Length 897;  
Best Local Similarity 21.2%; Pred. No. 0.26;  
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PYEVVARLP-----APA-KPEOD-----SDSDEGAAG-----PAGAPRTL 97  
DB 202 PMSLPALVPSPKRTWVSPAKAKYDEIFLTKDKMDGYSGLEVRETFKLTGLPSAL 261  
QY 98 VRRRRRLDPGEAVPVVYS-----GKVO-----SLNLIPNSSLK-----LC 138  
DB 262 LAH-----IWSLCDTRKGGCKLSKDOFALAFHLI--NQKLKIDPPHSLT 304  
QY 139 PS--EPEDADLTN--SGSSPED-----DALPGSPWRKKLRKKCKEKKKEEPPD 187  
DB 305 PEMIPSPDRSSLOKNITGSSPVADPSAIKELDTLNNEIYDLOREKNVNEODLKEKEDTVK 364  
QY 188 ODISPLPQSSSRKRTHEALQKLEVNKRLODRSCISPKOHQ-----SPALQSTDDE 242  
DB 365 QRTSEV--QDLQDEVQRESINLQKLAQKQVQELLGELDEKQAQLEQLODEVKKCAEE 422  
QY 243 VLVV--EGPVLPQSSRLFTLK---ICRADIVRLPVRMSEPLQNV-----VDHMANHL 290  
DB 423 AOLISLKAKEITSOESQISYEELKAREELSRLOQETAOLESEVESGKAQLEPLOOHL 482  
QY 291 GVSPPRIILLFGESELSPAT-----PSTLKLGVADIIDCVLASSSEATETSO 339  
DB 483 QESQOEISSMQRLEKDETDNOSNMSSSPQSVLVNGATDY--CSLSTSSSETANFNE 540  
QY 340 --ELRLVQKKEKHOMLEISLSPD--SPLKVLMSHYEAMGLSGH--KLSEFFDGTLSGKE 395  
DB 541 HAEQGNLSEPTHQESSVRSSPEIAPSDV--TDESEAVTAGNEKVTFRPDDKHSKEE 598  
QY 396 LPADLGLESGLIE 409  
DB 599 DP--FNVESSSLTD 610

RESULT 10  
US-08-477-389-4  
; Sequence 4, Application US/08477389  
; Patent No. 5872219  
; GENERAL INFORMATION:  
; APPLICANT: Diflore, Pier P  
; APPLICANT: Fazio, Francesca  
; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
; TITLE OF INVENTION: Factor Receptor Kinase  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,389  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/095,737  
; FILING DATE: 22-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelson, Ned A  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH060.001A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 897 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-477-389-4

Query Match 5.3%; Score 111; DB 2; Length 897;  
Best Local Similarity 21.2%; Pred. No. 0.26;  
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PYEVVARLP-----APA-KPEOD-----SDSDEGAAG-----PAGAPRTL 97  
DB 202 PMSLPALVPSPKRTWVSPAKAKYDEIFLTKDKMDGYSGLEVRETFKLTGLPSAL 261  
QY 98 VRRRRRLDPGEAVPVVYS-----GKVO-----SLNLIPNSSLK-----LC 138  
DB 262 LAH-----IWSLCDTRKGGCKLSKDOFALAFHLI--NQKLKIDPPHSLT 304  
QY 139 PS--EPEDADLTN--SGSSPED-----DALPGSPWRKKLRKKCKEKKKEEPPD 187  
DB 305 PEMIPSPDRSSLOKNITGSSPVADPSAIKELDTLNNEIYDLOREKNVNEODLKEKEDTVK 364  
QY 188 ODISPLPQSSSRKRTHEALQKLEVNKRLODRSCISPKOHQ-----SPALQSTDDE 242  
DB 365 QRTSEV--QDLQDEVQRESINLQKLAQKQVQELLGELDEKQAQLEQLODEVKKCAEE 422  
QY 243 VLVV--EGPVLPQSSRLFTLK---ICRADIVRLPVRMSEPLQNV-----VDHMANHL 290  
DB 423 AOLISLKAKEITSOESQISYEELKAREELSRLOQETAOLESEVESGKAQLEPLOOHL 482  
QY 291 GVSPPRIILLFGESELSPAT-----PSTLKLGVADIIDCVLASSSEATETSO 339  
DB 483 QESQOEISSMQRLEKDETDNOSNMSSSPQSVLVNGATDY--CSLSTSSSETANFNE 540  
QY 340 --ELRLVQKKEKHOMLEISLSPD--SPLKVLMSHYEAMGLSGH--KLSEFFDGTLSGKE 395  
DB 541 HAEQGNLSEPTHQESSVRSSPEIAPSDV--TDESEAVTAGNEKVTFRPDDKHSKEE 598  
QY 396 LPADLGLESGLIE 409  
DB 599 DP--FNVESSSLTD 610

RESULT 11  
US-09-061-709-2  
; Sequence 2, Application US/09061709B  
; Patent No. 629364  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Yao-Tseng  
; APPLICANT: Gure, Ali  
; APPLICANT: Tsang, Solam  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Jager, Elke  
; APPLICANT: Knuth, Alexander  
; APPLICANT: Old, Lloyd J.  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated  
; TITLE OF INVENTION: Antigens, The Antigens Per Se, And Uses Thereof  
; FILE REFERENCE: LUD 5538  
; CURRENT APPLICATION NUMBER: US/09/061,709B  
; FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 8

SEQ ID NO 2  
LENGTH: 1142  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-061-709-2

Query Match 5.2%; Score 109.5; DB 4; Length 1142;  
Best Local Similarity 23.0%; Pred. No. 0.49;  
Matches 98; Conservative 47; Mismatches 159; Indels 123; Gaps 20;

28 CRRARSPARL-IP-----DTVLVDLVN-DSDEVELEVADPEVP-----VARLPA 71  
25 CPCEGDSQSPLOIQSSPSDDDTLYLPQSPQNSDEGSDSDPIQRPEGKDSQSPLOIPQ 84  
QY 72 PAKPEDSDSDSGAAGPRGAPRTLYRRRRLLDPGEAPVVPVYSGKQSSLNLPD- 130  
DB 85 SSEGDDTQSPLONSQSSPEG-----KDSLPLEISQSPPEGDEDVQSPLONPASS 134  
QY 131 --NSLLKCPSEPE-----DEADLTNSGSSPSSEDDALPSGSPMRKRLRKCKE 177  
DB 135 FFSFALLSTFQSSPESTQSPFEGFPQSVQIPIVSAASSSTLVISFQSSP-----ES 185  
QY 178 EEKKMEFPDQDISPLPQSSRNKSRKHTALQKREVNKRLOD-----L 222  
DB 186 TQSPFEGFPQ---SPLOIPVSRFSSTLLSTFQSSPSRQRTSEGFQSPLOIPVSSSS 242  
QY 223 RGLSKQKQSPALQSTDEEVYVEGVPVLPQ-----SLFTLKIRCRADLYRLPYRMS 276  
DB 243 STLLSTFQSSPERTQST-----FEG-PPQSPLOIPVSRFSSTL-----LSIFQ 286  
QY 277 EP--LQNVVDHMAN--HLGVSPNRLLEGESELPTATPST-----LTKGVADI 323  
DB 287 SPERTQSTFEGFPQSPLOIPVSPSSSTLVISFQSSPERTQSTFEGFPQSPLOIPVSSSF 346  
QY 324 DCVYLA---SSSEATEPQSE---LRLRYQKKEKQMLEISL-----SPDS 362  
DB 347 SSTLLSTFQSSPERTQSTFEGFPQSPLOIPGSPFSSTLLSTFQSSPERTQSTFEGFPQ 406  
QY 363 PLKVLMS 369  
DB 407 PLOIPMT 413

RESULT 12  
US-08-955-918C-7  
Sequence 7, Application US/08955918C  
Patent No. 6268130  
GENERAL INFORMATION:  
APPLICANT: Kieyn, Patrick, and Moore, Karen  
TITLE OF INVENTION: RP Compositions and Therapeutic and  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/955,918C  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/697,766  
FILING DATE: 29-AUG-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MNT-007CPDV2C2PA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-955-918C-7

Query Match 5.2%; Score 108.5; DB 4; Length 504;  
Best Local Similarity 22.9%; Pred. No. 0.19;  
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY 3 EPLR-----GRGPRSRGGRARRARGRCRRAROSPAPRLPDTVLVDLVSDSEVL 56  
DB 44 EPLMVQANADGR-PRSR-----RAROSEQA---PLVESTLSSGSGTSTY 83  
QY 57 EVADPEVVPARLPAPAKPEDSDSDSGAAE--GPAGAPRTLYRRRRRLDPGEAPV 114  
DB 84 QVEADBSIASVQIGATRPAPASAKSKGAASGGGAGARKEKKHKTSGPA----- 138  
QY 115 PVTSGKQSSLNLPDNS-----SLKLCPEDEPADLTNSGSSPSSEDDALPSGSPMR 168  
DB 139 -----TLAEKSEAGQVQILTVGQSDHDKDGETAAGG-----AQPSGQDLR 182  
QY 169 KKLKCKEKEKKEEPPDDDISPLPQSSRNKSRKHTALQKREVNKRLODLRGLSP 228  
DB 183 ATWQKRGISSMFEDEDED-----ENSSSSQLMSNTPPSSANRSKSTREAAASP 236  
QY 229 KQKSPALQSTDEEVYVEGVPVLPQSSRLFTLKIRCRAD 267  
DB 237 AAEPP-----VDIEVQDLFEFALRPAPQGITIKRITRD 271

RESULT 13  
US-08-631-200-2  
Sequence 2, Application US/08631200  
Patent No. 5646040  
GENERAL INFORMATION:  
APPLICANT: Kieyn, Patrick W.  
APPLICANT: Moore, Karen J.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/631,200  
FILING DATE: 12-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-057  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864

```

;       TELEX: 66141 PENNIE
;       INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 505 amino acids
;           TYPE: amino acid
;           TOPOLOGY: unknown
;       MOLECULE TYPE: protein
;
US-08-631-200-2
```

Query Match	5.2%;	Score 108.5;	DB 1;	Length 505;
Best Local Similarity	22.9%;	Pred. NO. 0.19;		
Matches 64;	Conservative 36;	Mismatches 114;	Indels 65;	Gaps 10;

Qy	3	EPLR-----GGRSSSGGRRARRARGRCRPARSSPARLIPBVLVDVDSSEVL	56
Db	44	EPLWQANADR-PRSR-----RARSEDOA--PLVESYLSSGSTSY	83
Qy	57	EVADPEVPEVRLTAPAKPEQDSDSDSEGAEE--GPAAGEPILVRRRRRLPLGEPVY	114
Db	84	QVEADSTIASYOLGATRRPAPASAKKSGKAASGGCGAGPKKKKNGKTSIPA-----	138
Qy	115	PVYSGKVOSSLNLLPDPNS-----SLKLCIPSEPEDADLLNCGSSPSEDDALPSSGPMW	168
Db	139	-----TLAEKDSNAGPVQIILTVQSDHDKRAGETAAGG-----AGPSGDLR	182
Qy	169	KKLKKCKCEKKKEEPPDDOISLPDPSSNNKKRKTETALQIKREYNNKRLLODRGLSP	228
Db	183	ATMKQITGSSMSSTDEDED-----ENSSSSQLMSNTRPPSSATSRKSIETEAASAPSP	236
Qy	229	KOHOSPALQSTDDEVVLVEGEVLPPOSSRLPLTKTRCAD	267
Db	237	AAPEP-----VDIEVQDLEEFALPAPAGGITIKRITRD	271

US-08-630-592-4  
Sequence 4, Application US/08630592  
Patent No. 5770432  
GENERAL INFORMATION:  
APPLICANT: Nishina, patsy  
APPLICANT: No. 5770432entrauth, Konrad  
APPLICANT: Nagert, Juergen  
APPLICANT: No. 5770432th, Michael  
TITLE OF INVENTION: Obesity Associated Genes  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHER, HOEBACH, TEST, ALBRITTON & HERBERT  
STREET: 3400 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94114187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PCDOS/MSDOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,592  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J.  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: A59504/BIR/PUS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 7811989  
TELEFAX: (415) 3983249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 505 amino acids

```

;      TYPE: amino acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
;
US-08-630-592-4

```

Query Match	5.2%	Score 108.5;	DB 1;	Length 505;
Best Local Similarity	22.9%	Pred. No. 0.19;		
Matches 64; Conservative	36;	Mismatches 114;	Indels 65;	Gaps 10;

Qy	3	EP LR-----GRRGRSRRGRRARRARRGRCRCRPAKRSFARLLPDPVLYLVYSDSDSEVL	56
Db	44	EP LWOANADR-PRSR-----RAROSEDA---PLVESYLSSSGTSY	83
Qy	57	EVADPVEVYVRLRLAPAKPEQDSDSDSEGAEE--GPAGAPTVLRRRRRLPLPGEAPV	114
Db	84	QVEEDSTIASVQLTATRPAPAPASAKKSKGAASGGGAGAPRKEKKGKSTGPA-----	138
Qy	115	PVYSGKVQSSLNLLPQDNS-----SLKLQCEPEDEADLTNQSPPSEDDALPQSGPWR	166
Db	139	-----TLAEDKSEAGPVQIILLVQGDHDKBAGETAAAGG-----AQPSGD L R	182
Qy	169	KKLRKCKEKKKKMEFPDQDISLPQPSNNKRRKRTALQKRLVNNKRLQDLRCLSP	228
Db	183	ATMQKGTGSISSMSTDEDEDEDE-----ENSSSQSLMNSNTRPSSATSRKSTIREAASAP	236
Qy	229	KOHOSPALQSTDEDEVVLEGEVLPQSSRLFTLTKTRCAD	267
Db	237	AAPEPP-----VDIEVQDLEEFALPAPGQITIKCRITRD	271

RESULT 15  
 US-08-714-991-4  
 Sequence 4, Application US/08714991  
 Patent No. 576762  
 GENERAL INFORMATION:  
 APPLICANT: NORTH, Michael  
 APPLICANT: NISHINA, Patsy  
 APPLICANT: No. 5776782en-Trauth, Konrad  
 APPLICANT: NAGGERT, Juergen  
 TITLE OF INVENTION: OBESITY ASSOCIATED GENES  
 NUMBER OF SEQUENCES: 28  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
 STREET: 4 Embarradero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-4187  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/714,991  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: SHERWOOD, Pamela J.  
 REGISTRATION NUMBER: 36,677  
 REFERENCE/DOCKET NUMBER: A-59504-1/PJS  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-494-8700  
 TELEFAX: 415-494-8771  
 TELE: 910 277299  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 505 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-714-991-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;  
Best Local Similarity 22.9%; Pred. No. 0.19;  
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

QY	3	EPLE-----	GRGPRSRGGRARRARGRCPRAROSPARIIPDTVLYVDLVSDSEVL	56
DB	44	EPIMVQANADGR-PRSR-----	RAKQSEQA---PLVESYLSSSGSTSY	83
QY	57	EVADPVEVPVARIAPAPAKPEQDSDSSEGAEE--GPAGAPRTLVRRRRRLDPGEAPVY	114	
DB	84	QVQEAQDSIASVQIGATRRPPAPASAKKSGAAGGOGAPRKEKKGKGTSGPA-----	138	
QY	115	PVYSGKVQSSINLIPDNS-----	SLKLCPSPEDEADLTNGSSPSEDDALPSGSPWR	168
DB	139	-----	TLAEDKSEAQGPVQILTWGQSDHDKDAGETAAGG-----AQPSGODLR	182
QY	169	KLRLKCKEKEKEKEPPDODISPLPOPSSRNKSRKHTEALQKLRVKNRLODLRSGCLSP	228	
DB	183	ATMQRKGISMSMFDEDED-----	ENSSSSQLNSNTRPSSATSRKSIREAASAPSP	236
QY	229	KOHQSPALQSTDEVVYLVESPVLPQSSRLFTTKIRCRAD	267	
DB	237	AAPEPE---VDIEVQDLEEFALRPAPQGITIKCRITRD	271	

Search completed: April 22, 2002, 10:08:37  
Job time: 142 sec

This Page Blank (uspro)